Selbstorganisierende Empfehlungssysteme im Internet

Eine interdisziplinäre Betrachtung zwischen Technik und Recht

zur Erlangung des akademischen Grades eines

DOKTORS DER INGENIEURWISSENSCHAFTEN

von der Fakultät für Informatik
der Universität Fridericiana zu Karlsruhe (TH)

genehmigte

Dissertation

von

Dipl.-Inform.Wirt Christoph Sorge

aus Speyer

Tag der mündlichen Prüfung: 26. Oktober 2007

Erster Gutachter: Prof. Dr. Martina Zitterbart

Zweiter Gutachter: Prof. Dr. Thomas Dreier, M.C.J.

Vorwort

> *Certains auteurs, parlant de leurs ouvrages, disent: Mon livre, mon commentaire, mon histoire, etc. [...] Ils feraient mieux de dire: Notre livre, notre commentaire, notre histoire, etc., vu que d'ordinaire il y a plus en cela du bien d'autrui que du leur.*
>
> *Blaise Pascal*

Das Entstehen der vorliegenden Dissertation wäre ohne zahlreichende helfende Hände nicht möglich gewesen. Mein besonderer Dank gilt zunächst den beiden Gutachtern. Frau Professorin Dr. Martina Zitterbart als Erstgutachterin hat mich bei der Anfertigung der Arbeit nach Kräften unterstützt und beraten. Auch der Zweitgutachter, Herr Professor Dr. Thomas Dreier, M.C.J., war von Anfang an in die Betreuung der Arbeit eingebunden. Ich verdanke ihm wertvolle Ideen und Anregungen. Die Interdisziplinarität der vorliegenden Arbeit wurde erst durch die gute Zusammenarbeit mit beiden Gutachtern und das gegenseitige Verständnis für die jeweils andere Disziplin ermöglicht.

Bedanken möchte ich mich auch bei meinen Korrekturlesern. Frau Denise Dudek, Herr Marcel Noe und Herr Lars Völker haben in diese Tätigkeit viel Zeit investiert und so die Bereinigung so manchen Fehlers ermöglicht. Professor Dr. Jürgen Kühling, der eine frühere Version des Kapitels 3 korrekturgelesen hat, konnte im Bereich des Datenschutzrechts wertvolle Hinweise geben, die ich dankbar aufgenommen habe. Frau Denise Dudek und Herr Wenzel Svojanovsky haben mich darüber hinaus bei der Anfertigung diverser Abbildungen unterstützt.

Die vorliegende Dissertation entstand in einem Umfeld, das die interdisziplinäre Forschung in den Wirtschaftswissenschaften, der Informatik und der Rechtswissenschaft förderte. Neben der Finanzierung eines Großteils der Arbeit ergab sich im Graduiertenkolleg „Informationswirtschaft und Market Engineering“ auch die Gelegenheit, die Forschungsarbeit in diesen unterschiedlichen Bereichen kennenzulernen, Kontakte zu knüpfen und zu einer fruchtbaren Zusammenarbeit zu kommen.

Doch nicht nur das unmittelbare Arbeitsumfeld hat zum Zustandekommen dieser Dissertation beigetragen. Last, but not least, gilt mein besonderer Dank daher meiner Familie und meinen Freunden, die mich während meiner Promotion stets ermuntert und unterstützt haben.

Inhaltsverzeichnis

Abkürzungsverzeichnis

AG	Amtsgericht
AGB	Allgemeine Geschäftsbedingungen
Art.	Artikel
Az.	Aktenzeichen
B2C	Business to Consumer
BDDSG	Bundesdatenschutzgesetz
BGB	Bürgerliches Gesetzbuch
BGH	Bundesgerichtshof
BVerfG	Bundesverfassungsgericht
CA	Certification Authority (Zertifizierungsstelle)
DHT	Distributed Hashtable
DoS	Denial of Service
GbR	Gesellschaft bürgerlichen Rechts
GG	Grundgesetz für die Bundesrepublik Deutschland
GTIN	Global Trade Item Number
HTTP	Hypertext Transfer Protocol
IP	Internet Protocol
ISBN	International Standard Book Number
ISO	International Organization for Standardization
MD5	Message Digest 5
MDStV	Staatsvertrag über Mediendienste

NAT	Network Address Translation
NJW	Neue Juristische Wochenschrift
OSI	Open Systems Interconnection
PET	Privacy Enhancing Technologies
PKI	Public Key Infrastructure
Rn.	Randnummer
RStV	Rundfunkstaatsvertrag
Rz.	Randziffer
SHA-1	Secure Hash Algorithm 1
Sp.	Spalte
StGB	Strafgesetzbuch
StPO	Strafprozessordnung
TCP	Transmission Control Protocol
TDDSG	Teledienstedatenschutzgesetz
TDG	Gesetz über die Nutzung von Telediensten
TDSV	Telekommunikations-Datenschutzverordnung
TKG	Telekommunikationsgesetz
TKÜV	Telekommunikations-Überwachungsverordnung
UrhG	Gesetz über Urheberrecht und verwandte Schutzrechte (Urheberrechtsgesetz)
UWG	Gesetz gegen den unlauteren Wettbewerb

Abbildungsverzeichnis

Tabellenverzeichnis

Kapitel 1

Einleitung

1.1 Motivation

Der elektronische Handel hat in den letzten Jahren stark an Bedeutung gewonnen. So lag der Umsatz dieses Bereichs allein in Deutschland im Jahr 2005 bei 321 Milliarden Euro, davon immerhin 32 Milliarden Euro im Endkundengeschäft (Business-to-Consumer, B2C) [Bund06b]. Ein großes Manko des E-Commerce im Vergleich zu traditionellen Ladengeschäften ist jedoch der Mangel an persönlicher Beratung. Ein Verkäufer kann beispielsweise einem ihm bekannten Kunden Kaufempfehlungen geben, die er im Beratungsgespräch entwickelt oder aus der persönlichen Kenntnis der Interessen des Kunden ableitet. Ein Versuch, dieses Manko auszugleichen, liegt in der Entwicklung von *Empfehlungssystemen* [ReVa97], die dem Kunden personalisierte Bewertungen für ihm bekannte Produkte generieren oder auch Empfehlungen für vorher unbekannte Produkte aussprechen können. Erfolgreiche Unternehmen wie beispielsweise der Buchhändler Amazon (`http://www.amazon.de`, [LiSY03]) setzen solche Systeme bereits seit Jahren ein. Systeme eines einzelnen Händlers haben in der Regel aber den Nachteil, auf Produkte beschränkt zu sein, die dieser Händler auch anbietet. Übergreifende Empfehlungssysteme haben sich in Form von Portalen wie ciao.de und dooyoo.de in den letzten Jahren entwickelt. Gemein ist auch diesen Empfehlungsdiensten jedoch, dass sie von einem kommerziellen Betreiber auf einer zentralisierten Plattform betrieben werden. Für den Nutzer kann dies zu Nachteilen führen:

- Die zentrale Stelle ist aus technischer Sicht ein *Single Point of Failure*. Fällt sie aus oder ist sie nicht mehr erreichbar, so kann das Empfehlungssystem nicht mehr funktionieren. Die redundante Vorhaltung zentraler Komponenten kann dieses Problem teilweise ausgleichen, führt jedoch zu deutlich erhöhten Kosten.

- Auch unabhängig von möglichen technischen Fehlern ist das System vom Betreiber abhängig. So kann dieser zum Schutz eigener Interessen Inhalte beeinflussen oder filtern; wirtschaftliche Schwierigkeiten des Betreibers oder bewusste Managemententscheidungen können zum Verlust der gespeicherten Informationen führen.

Andere Anwendungen leiden oder litten unter ähnlichen Problemen. Der – oft rechtswidrige – Austausch urheberrechtlich geschützter Werke beispielsweise erfolgte lange Zeit nach dem Client-Server-Prinzip oder mit Hilfe von Systemen wie Napster, bei denen zwar der eigentliche Datenaustausch auf Peer-to-Peer-Basis funktioniert, aber dennoch zentrale Komponenten benötigt wurden. Die Robustheit des „Filesharings“ konnte in der Folge durch Dezentralisierung erheblich erhöht werden. Indem Netze wie Gnutella [Ripe01] und seine Nachfolger oder Bittorrent [PGES05] sich selbst organisieren, ohne auf eine zentrale Instanz angewiesen zu sein, wird erreicht, dass ein – juristischer oder technischer – Angriff auf einzelne Knoten nur geringe Auswirkungen auf das Netz als ganzes hat. Weil potentiell alle Netzteilnehmer gleichgestellt sind, spricht man auch von (reinen) Peer-to-Peer-Systemen.

Mit einiger Verzögerung folgten dem Filesharing weitere Anwendungen wie die internetbasierte Telefonie. Für diese wird beispielsweise mit dem Skype-System [BaSc04] ein – allerdings um zentrale Komponenten ergänztes – Peer-to-Peer-System eingesetzt. Auch zur Unterstützung des Electronic Commerce gibt es in der Forschung bereits Ansätze. Solche wurden beispielsweise im Rahmen des SESAM-Projekts an der Universität Karlsruhe (TH) entwickelt; genannt seien hier Peer-to-Peer-basierte Marktplattformen [CDHS+05] und Auktionsverfahren [RCNS06]. Sie ermöglichen es, auf einen einzelnen Betreiber zu verzichten und somit potentiell die Robustheit der erbrachten Dienstleistung zu erhöhen.

Es liegt nahe, bei der Konstruktion eines Empfehlungssystems, das ganz ähnliche Entwurfsziele erreichen soll, auf die gleichen Prinzipien zurückzugreifen. Die bestehenden Ansätze, die den elektronischen Handel auf Basis von Peer-to-Peer-Systemen unterstützen, können somit sinnvoll ergänzt werden. Konkret bedeutet dies, dass auf eine zentrale Instanz zur Berechnung von Empfehlungen verzichtet werden soll; auch die Datenhaltung soll dezentral erfolgen.

Ziel der vorliegenden Arbeit ist die Entwicklung eines solchen Systems, das Empfehlungen durch Einsatz eines Peer-to-Peer-Systems ohne zentrale Komponente ermöglicht. Auch, wenn die Motivation der Arbeit sich aus der Verwendung in Märkten ergibt, werden ökonomische Aspekte nur am Rande betrachtet. Die Herausforderungen der Entwicklung eines Empfehlungssystems liegen vielmehr im Bereich der Informatik.

Eine rein technische Betrachtung der Problemstellung reicht allerdings nicht aus. Zunächst gilt es bei der Entwicklung und dem Betrieb eines Systems, das Daten über die Präferenzen einer Person – also personenbezogene Daten – verwaltet, juristische Einschränkungen zu beachten. Wichtiger erscheint jedoch, dass das Recht auch helfen kann, die Erwartungen eines Nutzers an das System dort zu erfüllen, wo die Technik alleine dazu nicht in der Lage ist. Im Vergleich zu zentral funktionierenden Systemen ergeben sich durch den Einsatz von Peer-to-Peer-Systemen auch schwerwiegende neue juristische Problemstellungen. Dies betrifft beispielsweise die Verantwortlichkeit (insbesondere im Datenschutzrecht), aber

auch den Umgang mit Angriffen auf das System (Computerkriminalität). Auf einen Teil der auftretenden Fragestellungen will diese Dissertation eine Antwort geben.

1.2 Verwandte Arbeiten

Die vorliegende Arbeit nimmt bei der Betrachtung selbstorganisierender Empfehlungssysteme eine breite Perspektive ein; vergleichbare Arbeiten finden sich in der Literatur nicht. Zahlreiche Einzelprobleme, die in dieser Dissertation aufgegriffen werden, sind jedoch bereits diskutiert worden.

Aus technischer Sicht ist der Ausgangspunkt der Forschung an Empfehlungssystemen im Tapestry-System [GNOT92] zu sehen, das zur Filterung von E-Mails nach ihrer Relevanz entwickelt wurde. In der Folge wurden zahlreiche weitere verwandte Systeme entwickelt, von denen als prominenteste GroupLens [KMMH+97] für die Filterung von Usenet-Nachrichten sowie das durch den erfolgreichen praktischen Einsatz bekannt gewordene Amazon.com-Empfehlungssystem [LiSY03] zu nennen sind. Ein Überblick über entsprechende Arbeiten wird in [RZFK00] und [AdTu05] gegeben. Vertrauen als Grundlage von Empfehlungssystemen wird in diversen Arbeiten, darunter [O'Sm05] und [Golb06], thematisiert.

Als Ergänzung zentralisierter Empfehlungssysteme werden seit wenigen Jahren auch verteilte Systeme betrachtet. Insbesondere sei hier [Zieg05] genannt: In der Arbeit wird untersucht, wie der Einsatz von Taxonomien (also die Klassifikation von Objekten) zur Dezentralisierung von Empfehlungssystemen beitragen kann. Das in der Arbeit vorgeschlagene Empfehlungssystem basiert auf dem Einsatz dieser Taxonomien sowie auf vertrauensbasierten Ähnlichkeitsmaßen. Der Einsatz in Peer-to-Peer-Systemen, also unter Verzicht auf eine zentrale Kontrollinstanz, findet jedoch in [Zieg05] keine Beachtung. In [Kaza05] wird betrachtet, wie unter Verwendung von Web-Technologien verteilte Informationen herangezogen werden können, um Empfehlungen zu erzeugen; der Schwerpunkt dieser Arbeit liegt auf der Datenhaltung und -modellierung, die in der vorliegenden Dissertation lediglich am Rande diskutiert werden.

Erst vereinzelt befassen sich in der Literatur vorhandene Ansätze auch mit der Problematik der Empfehlungserzeugung in Peer-to-Peer-Systemen. Eine Zusammenfassung findet sich in Tabelle 1.1. Dabei ist u die Anzahl durch einen Nutzer bewerteter Objekte, t für nutzerbasierte Ansätze die Anzahl der betrachteten Nachbarschaft ähnlicher Knoten und n die Gesamtzahl an Knoten im Peer-to-Peer-System. Der Kommunikationsaufwand ergibt sich dabei aus der Anzahl auf Transportschicht versandter Nachrichten und bezieht sich jeweils auf das Einfügen einer einzelnen Bewertung bzw. die Berechnung einer einzelnen Empfehlung.

Die umfassendste Betrachtung findet sich in [MiKR04]; hier wird ein zunächst zentrales

Ansatz	Kommunikationsaufwand pro eingefügter Bewertung	Kommunikationsaufwand pro berechneter Empfehlung	Bemerkung
[PSWS05]	$O(t)$	$O(t)$	—
[MiKR04] UI-Chord	$O(t \cdot \log n)$	$O(t \cdot \log n)$	nicht implementiert
[MiKR04] II-Chord	$O(u \cdot \log n)$	$O(u \cdot \log n)$	nicht implementiert
[HXYS04]	$O(u \cdot \log n)$	$O(u \cdot \log n)$	durchschn. Aufwand durch Auswahl betrachteter Bewertungen reduziert; dadurch beschränkter Anwendungsbereich
[XHYS04]	$O(u \cdot \log n)$	$O(u \cdot \log n)$	Erweiterung von [HXYS04]
[OkMA04]	—	—	nur Architektur, keine Umsetzung
[WPLR06]	$O(u)$	$O(u)$	Anwendungsszenario unstrukturierte Peer-to-Peer-Filesharing-Systeme; dort geringerer zusätzlicher Kommunikationsaufwand

Tabelle 1.1: Peer-to-Peer-Empfehlungssysteme

System vorgestellt und verschiedene Ideen diskutiert, wie sich dieses auf ein Peer-to-Peer-System übertragen lässt. Hierzu zählt ein Ansatz (II-Chord), bei dem Daten über die gemeinsame Verwendung eines Objekts mit anderen Objekten in der verteilten Hashtabelle Chord [SMKK+01] unter dem Schlüssel der jeweiligen Objekte gespeichert werden. Dieser Ansatz wird von den Autoren nicht evaluiert; in der vorgestellten Form führt er beim Einfügen einer Bewertung zu einem linearen Kommunikationsaufwand in der Anzahl bis zu diesem Zeitpunkt abgegebener Bewertungen – sowohl beim Einfügen einer Bewertung als auch bei der Berechnung einer Empfehlung. Über die gesamte Laufzeit des Systems ergibt sich somit für das Einfügen von (insgesamt u) Bewertungen sogar ein quadratischer Aufwand; durch den Versand von Nachrichten im Chord-System erhöht dieser sich noch um den Faktor $\log n$ (vgl. Tabelle 1.1).

Ebenfalls in [MiKR04] wird ein Ansatz (UI-Chord) vorgeschlagen, bei dem als Schlüssel für die Speicherung von Bewertungen die Identität des jeweils Bewertenden verwendet wird. Jeder Knoten kann somit die Nutzerprofile einer Reihe anderer Knoten aus dem verwendeten Peer-to-Peer-System abrufen und daraus eigene Empfehlungen berechnen. Ein konkreter Algorithmus für das Auffinden dieser benachbarten Knoten fehlt ebenso wie eine Implementierung oder simulative Evaluierung. Der Ansatz bildet jedoch einen Ausgangspunkt für den nutzerbasierten Ansatz, wie er im Rahmen der vorliegenden Arbeit implementiert wird. Ein weiterer nutzerbasierter Ansatz, bei dem Bewertungen ähnlicher Knoten als Grundlage der Empfehlungsberechnung dienen, ist in [PSWS05] dargestellt. Da hier jeder Knoten nur mit seiner Nachbarschaft kommuniziert, hängt der Kommunikationsauf-

wand lediglich von der Größe dieser Nachbarschaft ab; auch hier erhöht sich der Aufwand durch den Versand von Nachrichten im Chord-System noch um den Faktor $\log n$.

In [OkMA04] wird eine Plattform vorgestellt, die Daten für verschiedene Empfehlungssysteme halten soll. Die Autoren postulieren die Anwendbarkeit in einem Peer-to-Peer-System (Chord [SMKK+01]), gehen jedoch nicht darauf ein, in welcher Weise dies konkret geschehen könnte. In [HXYS04] wird ein Peer-to-Peer-Empfehlungssystem vorgestellt, das darauf beruht, Nutzer zu finden, die Objekte identisch bewertet haben. Dafür wird als Schlüssel zum Auffinden von Bewertungen die Kombination aus Identität des bewerteten Objekts und dessen Bewertung verwendet. Vorteil dieses Verfahrens ist, dass ein Knoten leicht ähnliche Knoten finden kann: Wird eine Bewertung in das System eingefügt, kann der einfügende Knoten unter dem gleichen Schlüssel, den er zum Einfügen verwendet hat, auch Identitäten anderer Knoten finden, die das entsprechende Objekt gleich bewertet haben. Allerdings ist der Ansatz nur beschränkt anwendbar: Er kommt nur in Frage, wenn es lediglich eine kleine Menge möglicher Bewertungen gibt, da ansonsten die Wahrscheinlichkeit, dass überhaupt ein anderer Knoten ein Objekt gleich bewertet hat, zu gering ist. Zudem ist die Auswahl möglicher Ähnlichkeitsmaße beschränkt: Im allgemeinen Fall können Knoten sich ähnlich sein, ohne ein einziges Objekt genau gleich bewertet zu haben. Diese ähnlichen Knoten können mit dem Verfahren aus [HXYS04] jedoch nicht aufgefunden werden. In [XHYS04] wird dieser Ansatz ergänzt, indem eine Vorauswahl ähnlicher Knoten getroffen wird, deren Bewertungen für einen Knoten von Interesse sind. Beide Verfahren haben im schlechtesten Fall einen Kommunikationsaufwand entsprechend dem II-Chord-Verfahren. [WPLR06] beschreiben ein Empfehlungssystem für Filesharing-Netze, bei dem mit jeder Datei Informationen gespeichert werden, welche weiteren Dateien gemeinsam mit diesen Dateien genutzt wurden. Dieser Ansatz ist für das Erzeugen von Empfehlungen deshalb effizient, weil beim Download einer Datei die Information, welche anderen Dateien mit dieser gemeinsam genutzt wurden, sofort als Metainformation mit übertragen werden kann; es müssen also keine zusätzlichen Kommunikationsverbindungen aufgebaut werden. Der Ansatz setzt jedoch voraus, dass ein Knoten beim Download einer Datei andere Knoten, von denen er vorher Dateien bezogen hat, benachrichtigt, um die jeweiligen Profile zu aktualisieren. Für das Einfügen einer einzelnen Bewertung müssen also Nachrichten an u Knoten verschickt werden, wenn vorher bereits u andere Dateien heruntergeladen werden. Insgesamt entsteht (ähnlich [MiKR04]) über die Laufzeit des Systems so ein Kommunikationsaufwand, der mit der Anzahl der Dateien, die ein Nutzer herunterlädt, quadratisch wächst. Das System ist somit nicht beliebig skalierbar.

Gemein ist all diesen Ansätzen, dass sie sich nur auf einzelne Aspekte Peer-to-Peer-basierter Empfehlungssysteme konzentrieren. Die Betrachtung von Sicherheits- und Datenschutzaspekten tritt in den Hintergrund, die Skalierbarkeit der Systeme ist nur bei einem

Teil der Ansätze in Betracht gezogen, und simulative Evaluationen beschränken sich auf eine kleine Knotenzahl.

Aus juristischer Sicht findet sich bislang keine Literatur über Peer-to-Peer-basierte Empfehlungssysteme. Auch liegen keine Arbeiten vor, die sich spezifisch mit (zentralen) Empfehlungssystemen beschäftigen. In [Roßn07] sowie [Schw06] wird jedoch das Themengebiet der Personalisierung beleuchtet, in das sich auch Empfehlungssysteme einordnen lassen. Auch das verwandte Thema der Reputationssysteme hat – nicht zuletzt durch entsprechendes Interesse aus der Praxis, das durch den Erfolg des eBay-Reputationssystems entstanden ist – in die Literatur Eingang gefunden. Als umfassendste Darstellungen hervorzuheben sind hier [ScLa05] und [DöKo07].

Ebenfalls aus der Praxis heraus motiviert sind Beiträge, die sich mit juristischen Aspekten von Peer-to-Peer-Systemen beschäftigen. Diese gehen jedoch überwiegend von der Anwendung des Filesharing aus (so insbesondere [Wenz05, ReSc01]) und streifen Fragestellungen, die über das Urheberrecht hinausgehen, nur am Rande. Als einzige umfassende Darstellung jenseits des Urheberrechts kann [RaDH07] mit einer Betrachtung telekommunikationsrechtlicher Fragestellungen genannt werden. Die Autoren kommen zu dem Ergebnis, dass zumindest Teilfunktionalitäten von Peer-to-Peer-Systemen als Telekommunikationsdienste einzuordnen sind. Die Einordnung von Peer-to-Peer-Systemen aus Sicht des Telemedienrechts wird dabei jedoch ausdrücklich offen gelassen und wird auch sonst in der juristischen Literatur nicht betracht. Diese bildet deshalb auch den Schwerpunkt des juristischen Teils der vorliegenden Arbeit.

1.3 Gliederung

Das folgende Kapitel 2 gibt eine Einführung in Empfehlungssysteme und betrachtet die Grundlagen, die für das Verständnis der weiteren Kapitel benötigt werden. Datenschutz als wesentliche Anforderung für selbstorganisierende Empfehlungssysteme wird im dritten Kapitel betrachtet. Hierzu werden zunächst die Datenschutzanforderungen aus Nutzersicht betrachtet und in ein Modell aus fünf Ebenen eingeordnet; sodann wird die Rechtslage analysiert und diskutiert, inwieweit die bestehenden Datenschutzanforderungen durch das Recht erfüllt werden können. Kapitel 4 befasst sich mit dem Phänomen Vertrauen und seiner Behandlung in Rechtswissenschaft und Informatik. Es bildet somit eine wichtige Grundlage für den eigentlichen Entwurf eines Empfehlungssystems, der in Kapitel 5 behandelt wird. Es zeigt sich, dass die Betrachtung von Vertrauensbeziehungen sowohl die eigentliche Funktionalität des Systems unterstützen als auch eine missbräuchliche Verwendung erschweren kann.

Ob und inwieweit die angestrebten Ziele erreicht werden können, wird in Kapitel 6 untersucht. Dazu werden zunächst Metriken betrachtet, mit deren Hilfe die Bewertungsqualität

evaluiert werden kann; sodann wird – neben dieser qualitativen Betrachtung – die Skalierbarkeit, Robustheit und Sicherheit des Systems analysiert. Auch die im System gesetzten Anreize und die Erfüllung juristischer Vorgaben werden berücksichtigt.

Den Abschluss der Arbeit bietet ein Ausblick aus Sicht der Rechtswissenschaft: Probleme der derzeitigen Rechtslage im Bereich selbstorganisierender Empfehlungssysteme werden zusammengefasst sowie Empfehlungen für eine Fortentwicklung des Rechts gegeben.

Kapitel 2

Grundlagen

Dieses Kapitel beschreibt wesentliche Grundlagen, die für den Entwurf eines Empfehlungssystems benötigt werden.

2.1 Grundlagen aus Sicht der Telematik

Aus Sicht der Telematik ist ein selbstorganisierendes Empfehlungssystem zunächst eine Anwendung der Telekommunikation. An dieser Stelle werden deshalb Grundlagen von Kommunikationssystemen betrachtet: Zum einen die Strukturierung solcher Systeme in Schichten, zum anderen die Funktionsweise von Peer-to-Peer-Systemen als selbstorganisierende Kommunikationssysteme.

2.1.1 Schichtenmodelle der Telekommunikation

Protokolle, die zur Kommunikation innerhalb von Rechnernetzen verwendet werden, lassen sich in ein Schichtenmodell einordnen. Die Strukturierung anhand eines solchen Modells hat den Vorteil, dass bei der Betrachtung einer spezifischen Aufgabenstellung von (internen) Details darunter- und darüberliegender Schichten abstrahiert werden kann. Zu diesem Zweck bietet jede Schicht der darüberliegenden einen Dienstzugangspunkt (Service Access Point, SAP) an und verwendet den Dienstzugangspunkt der darunterliegenden Schicht, um deren Dienste in Anspruch nehmen zu können.

In der Telekommunikation wird gedanklich meist das sogenannte ISO/OSI-Schichtenmodell (für ISO/Open Systems Interconnection) [Inte94] zugrunde gelegt, das von der International Organization for Standardization standardisiert wurde. In diesem Modell werden sieben Schichten unterschieden (für eine ausführlichere Darstellung siehe [Zimm80]):

- Die *Bitübertragungsschicht* (Schicht 1), die physikalisches Übertragungsmedium und Übertragungsverfahren spezifiziert. Behandelt wird in dieser Schicht die Übertragung von Bits zwischen zwei benachbarten Stationen.

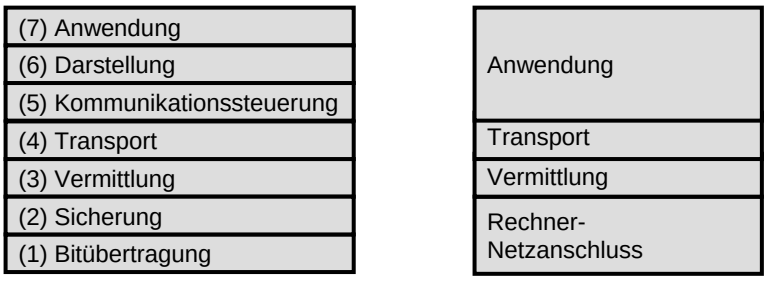

Abbildung 2.1: ISO/OSI- und TCP/IP-Schichtenmodell gegenübergestellt

- Die *Sicherungsschicht* (Schicht 2), die die Übertragung von Dateneinheiten zwischen zwei benachbarten Stationen ermöglicht. Fehlererkennung und -behebung können, falls benötigt, in dieser Schicht realisiert werden.

- Die *Vermittlungsschicht* (Schicht 3), die für die Wegewahl und Weiterleitung von Paketen zuständig ist. Erst sie ermöglicht die weltweite Kommunikation zwischen Kommunikationspartnern. In dieser Schicht ist auch die Adressierung der beteiligten Stationen angesiedelt.

- Die *Transportschicht* (Schicht 4), die den (zuverlässigen oder unzuverlässigen) Datentransport zwischen Endsystemen (Ende-zu-Ende-Transport) ermöglicht. Sie stellt der Kommunikationssteuerungsschicht eine Schnittstelle zum Transport von Nachrichten weitgehend unabhängig von den Chrakteristika des darunter liegenden Netzes zur Verfügung [Hals96].

- Die *Kommunikationssteuerungsschicht* (Schicht 5), die für die Verwaltung von Kommunikationssitzungen zuständig ist. Sie stellt auch Synchronisationspunkte zur Verfügung, an denen eine zwischenzeitlich unterbrochene Kommunikation wieder aufgenommen werden kann.

- Die *Darstellungsschicht* (Schicht 6), die eine systemunabhängige Darstellung von Daten erreicht.

- Die *Anwendungsschicht* (Schicht 7), die Anwendungen eine Schnittstelle zur Kommunikation zur Verfügung stellt und dabei vielfältige Funktionalitäten unterstützen kann.

Wenn auch das ISO/OSI-Schichtenmodell als Referenz große Beachtung gefunden hat und noch findet, ist – zumindest bezogen auf das Internet – in der Praxis das TCP/IP-Modell [RePo87] von größerer Bedeutung. Abbildung 2.1 stellt die beiden Schichtenmodelle gegenüber. Die untersten beiden sowie die obersten drei Schichten des ISO/OSI-Schichtenmodells sind hier jeweils in einer Schicht zusammengefasst: Bitübertragungs- und Sicherungsschicht

werden zur Rechner-Netzanschluss-Schicht, Kommunikationssteuerungs-, Darstellungs- und Anwendungsschicht zur Anwendungsschicht zusammengefasst.

2.1.2 Sichere Kommunikation

Der Einsatz kryptographischer Verfahren dient u.a. der Gewährleistung von Datensicherheit und Datenschutz, wie sie für das im Rahmen dieser Arbeit zu entwickelnde System unterstützt werden sollen. Im Folgenden werden deshalb Anforderungen an die sichere Übermittlung von Nachrichten aufgezeigt und ihre Umsetzung durch kryptographische Verfahren skizziert.

2.1.2.1 Anforderungen an die sichere Nachrichtenübertragung

Als wesentliche Anforderungen an die sichere Übertragung von Nachrichten lassen sich identifizieren (vgl. [TrWa02, S. 9]):

- *Integrität:* Eine Nachricht kann während des Transports nicht unbemerkt verändert werden.
- *Authentizität:* Der Empfänger kann sich der Identität des Urhebers der Nachricht sicher sein.
- *Nicht-Abstreitbarkeit:* Der Absender kann nicht mit Erfolg die Integrität und Authentizität einer von ihm stammenden Nachricht bestreiten.[1]
- *Vertraulichkeit:* Kein Dritter soll vom Inhalt einer Nachricht Kenntnis nehmen können.

2.1.2.2 Symmetrische und Asymmetrische Kryptographie

Kryptographie ist die Wissenschaft der Verschlüsselung von Informationen. Heute übliche Verschlüsselungsverfahren können als Funktion $f(m,k) = c$ beschrieben werden, wobei m die Originalnachricht, c das Chiffrat (die verschlüsselte Nachricht), k einen Schlüssel und f eine öffentlich bekannte Verschlüsselungsfunktion bezeichnen. Verfahren, bei denen für die Entschlüsselung einer Nachricht m derselbe Schlüssel k wie für die Verschlüsselung verwendet wird, werden als *symmetrische* Verfahren bezeichnet. Will ein Nutzer (Alice) einem anderen Nutzer (Bob) eine Nachricht zusenden, die vertraulich bleiben soll, so benötigen sie bei einem solchen Verfahren einen Schlüssel, der Alice und Bob (aber keinem Dritten) bekannt ist.

Bei *asymmetrischen* Verfahren (Public-Key-Kryptographie) werden zwei verschiedene Schlüssel verwendet. Dies sind der (geheim zu haltende) private Schlüssel e und der öffentliche Schlüssel d. Beide hängen so zusammen, dass eine mit dem öffentlichen Schlüssel

[1] In der Literatur wird unter den Begriff der Nicht-Abstreitbarkeit zum Teil auch gefasst, dass der Empfänger den Erhalt einer Nachricht nicht bestreiten können soll.

verschlüsselte Nachricht mit dem privaten wieder entschlüsselt werden kann.[2] Wichtig für die Funktionsweise ist, dass der private Schlüssel aus dem öffentlichen Schlüssel nur mit unverhältnismäßig hohem Aufwand zu berechnen ist. Will Alice Bob eine Nachricht zusenden, die vertraulich bleiben soll, so benötigt Alice dazu Bobs öffentlichen Schlüssel; Bob kann die Nachricht mit seinem privaten Schlüssel entschlüsseln. Wohl bekanntestes Beispiel eines solchen Verfahrens ist der RSA-Algorithmus, vorgeschlagen in [RiSA78].

2.1.2.3 Elektronische Signaturen

Neben der reinen Verschlüsselung, die dem Schutz der Vertraulichkeit dient, können kryptographische Verfahren auch zur Gewährleistung von Integrität, Authentizität und Nicht-Abstreitbarkeit eingesetzt werden. Ein gängiger Mechanismus zur Erreichung dieser Ziele ist die elektronische Signatur, zu deren Erzeugung in der Regel kryptographische Hash-Funktionen eingesetzt werden. Eine *Hash-Funktion* ist definiert als eine leicht zu berechnende Funktion, die eine Eingabe beliebiger Länge auf eine Ausgabe fester Länge abbildet [Wätj03, S.87].

Kryptographische Hash-Funktionen sind Hash-Funktionen, die den folgenden Anforderungen genügen (vgl. [TrWa02, S. 182], [Wätj03, S.87-90]):

- Es ist nicht effizient möglich, zu einem gegebenen Funktionswert (Hash-Wert) ein Urbild zu bestimmen.
- Es ist nicht effizient möglich, zu einer gegebenen Nachricht eine zweite zu finden, der der gleiche Funktionswert zugeordnet wird (*schwache Kollisionsresistenz*).
- Es ist nicht effizient möglich, zwei Nachrichten zu konstruieren, die auf den gleichen Funktionswert abgebildet werden (*starke Kollisionsresistenz*).

Will ein Nutzer (Alice) nun eine Nachricht m signieren, so berechnet er zunächst deren Hash-Wert, wendet eine Funktion darauf an, in die ihr privater Schlüssel eingeht – diese soll als Verschlüsselung bezeichnet werden – und fügt das Ergebnis der Nachricht hinzu. Will ein Empfänger (Bob) die Signatur prüfen, so berechnet er ebenfalls den Hash-Wert von m. Danach entschlüsselt er den Hash-Wert, der der Nachricht beigefügt war, mit Alices öffentlichem Schlüssel d. Stimmen beide Werte überein, so sind Authentizität und Integrität der Nachricht sichergestellt. Darüber hinaus kann Alice nicht mehr erfolgreich abstreiten, die Nachricht signiert zu haben. Der Signaturvorgang ist in Abbildung 2.2 dargestellt.

Gängige kryptographische Hashfunktionen sind z.B. MD5 (128 bit Ausgabe, dokumentiert in [Rive92]) und SHA-1 (160 bit Ausgabe, dokumentiert in [SHA195]). In jüngerer Zeit wurden allerdings Angriffe auf weit verbreitete Hashfunktionen, insbesondere MD5 [WaYu05], veröffentlicht.

[2] Wenn dies in beide Richtungen funktioniert, ist es möglich, mit nur einem Schlüsselpaar Nachrichten zu signieren (vgl. Abschnitt 2.1.2.3) und zu verschlüsseln.

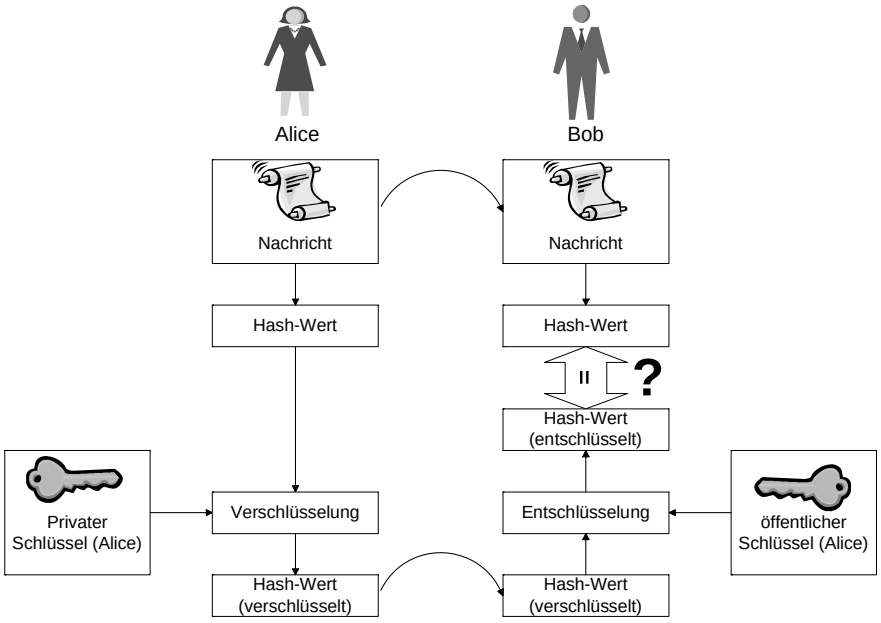

Abbildung 2.2: Elektronische Signatur [Sorg05]

2.1.2.4 Zertifizierung

Ein verbleibendes Problem des Ansatzes der elektronischen Signatur ist die sichere Herstellung der Zuordnung zwischen einer Person bzw. einer Identität und einem öffentlichen Schlüssel. Damit nicht jeder, der die elektronische Signatur eines Nutzers prüfen will, den öffentlichen Schlüssel persönlich von diesem Nutzer erhalten muss, wird das Prinzip der *Zertifizierung* angewendet. Jeder Nutzer besorgt sich einmalig den öffentlichen Schlüssel einer vertrauenswürdigen Instanz, der Zertifizierungsstelle (engl. certification authority, CA). Diese wiederum überprüft die Identität von Nutzern, die sicher kommunizieren wollen, und stellt diesen ein Zertifikat (ID-Zertifikat, Identitätszertifikat) aus. Das Zertifikat besteht aus einem öffentlichen Schlüssel, Daten, die diesem Schlüssel zugeordnet werden sollen, und der Signatur der vorgenannten Werte durch die Zertifizierungsstelle. Mit dem öffentlichen Schlüssel der CA kann jedes von ihr ausgestellte Zertifikat überprüft werden. Statt nur einer Zertifizierungsstelle können auch mehrere (nebeneinander oder in einer Hierarchie) bestehen.

Die Infrastruktur, die zum Management von Identitätszertifikaten benötigt wird und die Authentifizierung der Bindung eines öffentlichen Schlüssels an eine Identität ermöglicht, heißt *Public Key Infrastructure (PKI)* [BMBC+05].

2.1.3 Selbstorganisation

Im Rahmen der vorliegenden Arbeit werden *selbstorganisierende* Empfehlungssysteme betrachtet. In Anlehnung an [PrBe05] wird dabei die folgende Definition der Selbstorganisation zugrunde gelegt:

Definition 1. Ein System aus mehreren Komponenten ist selbstorganisierend, wenn

- eine lokale Interaktion zwischen Komponenten des Systems besteht,
- sich durch diese Interaktion ein emergentes Verhalten des Systems ergibt, also auf der Makroebene eine Funktionalität entsteht, die die einzelnen Komponenten für sich genommen nicht zur Verfügung stellen,
- sich das System an Änderungen in der Umgebung anpassen kann und
- diese Eigenschaften nicht durch Eingriffe von außerhalb des Systems entstehen.

Die vorliegende Arbeit befasst sich dabei ausschließlich mit einem verteilten, selbstorganisierenden System, das ohne zentrale Komponenten auskommt und keiner einheitlichen Verwaltung oder Kontrolle unterliegt.

2.1.4 Overlay-Netze

Mit dem Begriff *Overlay-Netz* soll in dieser Arbeit – der Definition aus [CGBD+05] folgend – ein logisches Kommunikationsnetz bezeichnet werden, das auf einem oder mehreren bestehenden Netzen (den Underlays) aufsetzt, um die Heterogenität dieser Netze zu verstecken, zusätzliche Dienste zu ermöglichen oder eine hohe Adaptivität oder Konfigurierbarkeit zu erreichen. So kann beispielsweise eine zusätzliche Adressierung hinzugefügt werden, wie dies bei Peer-to-Peer-Overlay-Netzen oft der Fall ist.

2.1.5 Peer-to-Peer

Grundidee von Peer-to-Peer-Systemen ist die gleichberechtigte Kommunikation zwischen Knoten – ohne herausgehobene Rolle einzelner Kommunikationspartner. Peer-to-Peer-Systeme sind kein neues Konzept – grundsätzlich waren im Internet von seiner Entstehung an die beteiligten Endsysteme gleichberechtigt.

Routing-Protokolle (wie OSPF [Moy98]) können als prominentestes Beispiel für den Einsatz des Peer-to-Peer-Prinzips genannt werden genannt werden; doch auch auf Anwendungsschicht existieren schon seit langer Zeit Protokolle, bei denen die beteiligten Knoten gleichberechtigt sind – beispielsweise für den Austausch von Nachrichten, die in den Diskussionsforen des Usenet veröffentlicht werden [RFC1036]. In diesem Fall sind die gleichberechtigten Knoten allerdings nicht jeweils nur einem Nutzer zugeordnet, sondern bieten ihrerseits in der Regel einen Dienst für viele Nutzer an.

Dennoch ist für viele Anwendungsbereiche das Client/Server-Modell vorherrschend. Hier wird zwischen einem Dienstgeber (Server) und einem Dienstnehmer (Client) differenziert, die unterschiedliche Aufgaben wahrnehmen. Als Beispiel kann der Abruf einer Website dienen: Der Web-Browser als Clientanwendung sendet eine Anfrage an den Server, der die Website daraufhin ausliefert. Dazu wird das HTTP-Protokoll [RFC2616] verwendet, in dem die Rollen von Dienstnehmer und Dienstgeber klar unterschieden werden.

Bei Peer-to-Peer-Systemen kann grundsätzlich jeder Teilnehmer beide Rollen einnehmen. Zwar gibt es Peer-to-Peer-Systeme, in denen für bestimmte Aufgaben (wie das Bereithalten eines Index oder Aufgaben der Authentifizierung und ggf. Abrechnung) ein zentraler Ansatz gewählt wird; in der vorliegenden Arbeit soll aber von sogenannten *reinen* Peer-to-Peer-Systemen ausgegangen werden. Es wird also auf jegliche zentrale Komponenten verzichtet.

In Anlehnung an [StWe04][3] werden (reine) Peer-to-Peer-Systeme wie folgt definiert:

Definition 2. Ein Peer-to-Peer-System ist ein selbstorganisierendes System, dessen Komponenten gleichberechtigte und autonome Einheiten sind, die einander Ressourcen zur Verfügung stellen und die nicht auf zentrale Dienste zurückgreifen.

Als Ressourcen kommen dabei beispielsweise Rechenleistung oder Speicherplatz in Frage.

Die in der vorliegenden Arbeit betrachteten Peer-to-Peer-Systeme lassen sich untergliedern in ein Overlay-Netz und eine darauf aufbauende Anwendung, die ihrerseits wiederum in mehrere Schichten gegliedert sein kann. Diese Modellierung erlaubt es, die Anwendung soweit wie möglich unabhängig von der Overlay-Implementierung zu gestalten und gegebenenfalls das verwendete Overlay-Netz auch auszutauschen, ohne dass dabei wesentliche Änderungen der Anwendung erforderlich werden.

Peer-to-Peer-Systeme lassen sich in zwei Kategorien einteilen: Strukturierte und unstrukturierte Systeme [DuGH03].

2.1.5.1 Unstrukturierte Peer-to-Peer-Systeme

Unstrukturierte Systeme zeichnen sich insbesondere dadurch aus, dass ein teilnehmender Knoten kein Wissen darüber hat, wo er welche Ressourcen im Netz auffinden kann [DuGH03]. Dies führt zu einer ineffizienten Suche nach Ressourcen, andererseits aber auch einer hohen Robustheit. Typischer Vertreter dieser Klasse von Peer-to-Peer-Systemen ist Gnutella [Ripe01].

[3] Die Autoren sehen ihre Definition wiederum als Verfeinerung der Überlegungen in [Oram01]

2.1.5.2 Strukturierte Peer-to-Peer-Systeme

Bei strukturierten Systemen hilft lokales Wissen einem Knoten bei der Suche nach Ressourcen [DuGH03]; beispielsweise weiß er, an welchen Knoten er eine Anfrage senden muss, um möglichst schnell eine bestimmte Ressource zu erreichen. Dies ermöglicht eine effizientere Suche, erfordert jedoch auch Aufwand, um die lokale Information aktuell zu halten.

Als konkretes Beispiel soll das Chord-Protokoll [SMKK+01] dienen. Das Verfahren basiert auf einer verteilten Hashtabelle (Distributed Hashtable, DHT): Auf den Identifikator (beispielsweise die IP-Adresse) eines jeden Knoten wird eine Hashfunktion angewendet; so ergibt sich sein Chord-Identifikator – ein Zahlenwert, der als Adresse des Knotens im Overlay-Netz fungiert. Einer Ressource, die im Netz vorhanden ist, wird ebenfalls ein Identifikator (beispielsweise ein Dateiname) zugeordnet. Dieser Identifikator heißt auch *Schlüssel*. Auf den Identifikator wird die gleiche Hash-Funktion angewendet wie auf den Identifikator eines Knoten. So ergibt sich, wo in der DHT die Ressource aufgefunden werden soll: Ein Knoten ist dann für einen Bereich der DHT zuständig, der in der Nähe des eigenen Identifikators liegt.

Im Fall von Chord lässt sich die DHT als Ring darstellen. Der Umfang des Rings entspricht dabei der Anzahl möglicher Ausgabewerte der verwendeten Hash-Funktion. Ein Knoten ist zuständig für alle Ressourcen, die im Ring (bei Durchqueren im Uhrzeigersinn) zwischen dem Chord-Identifikator seines Vorgängers im Ring und seinem eigenen liegen.

Jeder Knoten kennt nun, neben der Adresse seines Vorgängers, $O(\log N)$ (N Anzahl der Netzteilnehmer) weitere Knoten im Netz; er kennt also zu einer bestimmten Position im Ring jeweils die IP-Adresse des Knotens, dessen Chord-Identifikator dieser Position entspricht. Die Datenstruktur, in der diese Zuordnung abgelegt ist, heißt *finger table*. Der Knoten kennt dabei viele Knoten in seiner näherem Umgebung auf dem Chord-Ring und wenige in weiter entfernten Regionen des Rings. Insbesondere ist in der finger table die IP-Adresse des direkten Nachfolgerknotens enthalten. Die Knoten, die in der finger table enthalten sind, können auch als *Nachbarn* bezeichnet werden. Die Speicherung von $O(\log N)$ Knoten in der finger table führt dazu, dass auch beim Routen einer Nachricht zum Zielknoten im Chord-Ring im Durchschnitt $O(\log N)$ Schritte (Hops) benötigt werden.

Abbildung 2.3 zeigt ein vereinfachtes Beispiel des Auffindens einer Ressource: Knoten 11 kennt die Knoten, die für die Ressourcen 12, 13, 18, 15 und 3 zuständig sind.[4] Er möchte nun auf die Ressource Nummer 1 zugreifen und kontaktiert deshalb den ihm bekannten Knoten mit dem größten Chord-Identifikator, der kleiner ist als 1 (im vorliegenden Fall modulo 20 gerechnet, da der Chord-Ring einen Umfang von 20 hat). Es handelt sich hier um Knoten 18 (Abbildung 2.3 a). Die finger table von Knoten 18 hat die Einträge 19, 0, 2, 5 und 10. Der nächste betrachtete Knoten ist nun Knoten 0, da dieser den größten Chord-Identifikator

[4] Zur Vereinfachung wurde in der Grafik angenommen, dass zu jeder adressierbaren Ressource tatsächlich genau ein Knoten gehört. Dies ist in einem echten Chord-Netz nicht realistisch.

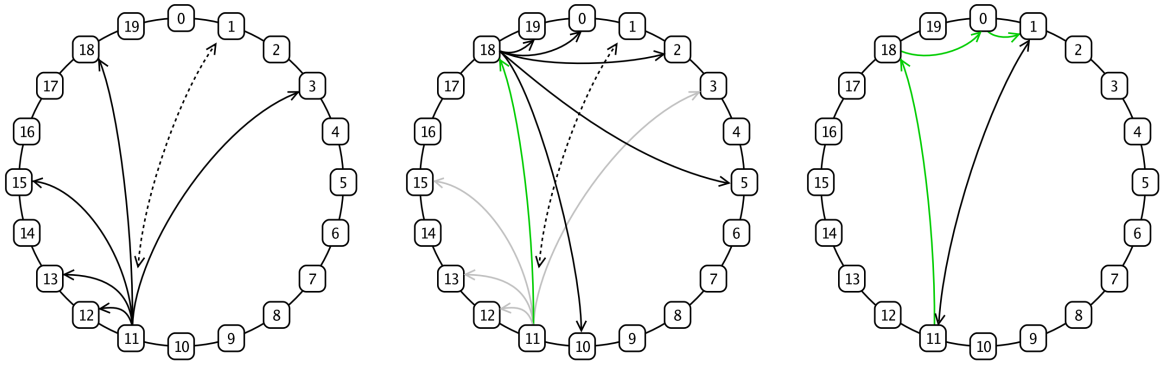

Abbildung 2.3: Auffinden einer Ressource mit Chord

der Knoten aus der finger table von Knoten 18 hat, der kleiner ist als 1 (Abbildung 2.3 b). Im Fall von iterativen Lookup-Operationen gibt Knoten 18 diese Information an Knoten 11 weiter, der dann Knoten 0 kontaktiert; bei rekursivem Lookup leitet Knoten 18 die Anfrage von Knoten 11 selbst an Knoten 0 weiter. Dieser wiederum kennt direkt den gewünschten Zielknoten (Abbildung 2.3 c).

Das Chord-Protokoll [SMKK+01] selbst spezifiziert keine Möglichkeit zur Speicherung von Daten, sondern lediglich das Routing zu einem zuständigen Knoten. Wird ein auf Chord basierendes Peer-to-Peer-System zur Speicherung von Dokumenten oder anderen Daten eingesetzt, ist also eine Ergänzung der Spezifikation erforderlich. Im einfachen Fall wird ein Dokument stets auf dem Knoten gespeichert, der nach dem Chord-Protokoll für den jeweiligen Schlüssel zuständig ist. Bei Ausfall dieses Knotens oder falls er das Netz regulär verlässt, ist dies alleine jedoch nicht ausreichend.

In [DBKK+01] wird eine einfache Ergänzung des Protokolls, das DHASH-System, vorgeschlagen: Ein eingefügtes Dokument wird nicht nur auf einem Nachfolgerknoten der Chord-ID des Inhalts, sondern auf den r folgenden Knoten gespeichert. Auf diese Weise wird die Zuverlässigkeit der Speicherung erhöht. Zudem ist die Übertragung von Dokumenten auf neu hinzukommende Knoten vorgesehen. [DBKK+01] spezifiziert weiterhin einen Cache-Mechanismus zur Erhöhung der Zugriffsgeschwindigkeiten im System. Wird ein Schlüssel über mehrere Hops hinweg gesucht, wird das zugehörige Dokument im Anschluss auf jedem Knoten auf dem Pfad gespeichert. Dies führt zu einer Verteilung insbesondere besonders häufig nachgefragter Dokumente.

2.1.6 Betrachtete Systembestandteile

Im Rahmen des entwickelten Systems werden die folgenden Akteure und Komponenten betrachtet:

- Ein *Nutzer* ist eine natürliche Person, die mit dem Empfehlungssystem interagiert.
- Ein *Knoten* ist die einem einzelnen Nutzer zugeordnete Systemkomponente. Das Peer-to-Peer-System besteht aus vielen solcher Knoten.

Nutzer und Knoten gemeinsam werden als *Teilnehmer* des Systems bezeichnet.

2.1.7 Angriffe auf Peer-to-Peer-Systeme

In der Literatur ist eine Reihe von Angriffen auf Peer-to-Peer-Systeme beschrieben, die es auch in der vorliegenden Arbeit zu berücksichtigen gilt. Als bekanntester Angriff ist der Sybil-Angriff zu nennen. Hierbei tritt ein Angreifer unter mehr als einer Identität im Netz auf. Verlässt sich ein Protokoll darauf, dass nur eine kleine Anzahl bösartiger Knoten am Netz teilnimmt, kann der Angriff die Funktionalität erheblich beeinträchtigen. In [Douc02] wird gezeigt, dass eine ausschließlich auf dezentralen Mechanismen basierende Abwehr des Sybil-Angriffs unter realistischen Annahmen nicht möglich ist. Es bestehen aber zwei Möglichkeiten, dennoch Abhilfe zu schaffen.

- Zum einen kann das Problem durch Verwendung einer zentralen Instanz gelöst werden. Denkbar sind beispielsweise die zentralisierte Vergabe der IDs [Wall03], [FrRe01] oder der Rückgriff auf ohnehin vorhandene – durch zentrale Stellen vergebene – externe Identifikatoren, wie beispielsweise digitale Zertifikate.
- Zum anderen kann der Sybil-Angriff durch dezentral funktionierende Mechanismen zwar nicht verhindert, aber doch erschwert werden. Denkbar ist die Verwendung sogenannter *Crypto Puzzles*, bei der ein beitretender Knoten beweisen muss, eine Rechenaufgabe gelöst zu haben (vgl. dazu Abschnitt 5.5.2.3, in dem diese Idee erneut aufgegriffen wird), oder der in [DiHa06] vorgeschlagene Prozess der „Selbstregistrierung“, der auf die IP-Adressen der Knoten zurückgreift und eine Überprüfung der so erlangten ID durch andere Teilnehmer des Netzes vorsieht.

Es ist davon auszugehen, dass die vorgestellten Lösungsansätze hinreichend sind, um den Sybil-Angriff für die Einnahme einer großen Anzahl an Identitäten zumindest so teuer zu gestalten, dass er sich für den Zweck der Manipulation eines Empfehlungssystems nicht lohnt. Als besondere Anforderung an das Verfahren zur Vergabe der Identitäten ist jedoch zu erwähnen, dass langlebige Identitäten erforderlich sind. Haben die Teilnehmer an dem Netz dynamisch zugewiesene IP-Adressen, kann der in [SMKK+01] beschriebene Mechanismus, die Knoten-ID als Hash-Wert der IP-Adresse zu ermitteln, diese nicht sicherstellen. Die anderen, oben erwähnten Verfahren haben dieses Problem jedoch nicht.

Er bleibt jedoch eine Reihe weiterer möglicher Angriffe gegen Overlay-Netze, die an dieser Stelle betrachtet werden müssen. Im Bereich des Routing sind zu nennen:

- der Eclipse-Angriff. Hierbei werden Knoten geschickt so im Netz positioniert, dass sie die durch Verzicht auf das Weiterleiten von Nachrichten bestimmte andere Knoten von der Kommunikation abschneiden können. Bei Chord ist durch seine strukturellen Eigenschaften – insbesondere die Tatsache, dass jeder Knoten sowohl Knoten in seiner unmittelbaren Umgebung als auch in weit entfernten Bereichen des Chord-Rings kennt – dieser Angriff jedoch sehr aufwendig [SCDR04].

- der Churn-Angriff, bei dem zahlreiche Knoten in rascher Folge dem Netz beitreten und es wieder verlassen. Da beim Chord-Protokoll für Beitritt und Austritt eines Knotens jeweils $O((\log N)^2)$ Nachrichten ausgetauscht werden müssen [SMKK+01], stellt dies eine höhere Netzbelastung dar als einfache Lookup-Operationen. Auch dieser Angriff erfordert jedoch die Kontrolle über zahlreiche Knoten. Der Umgang mit Churn ist nicht Gegenstand der vorliegenden Arbeit; für eine ausführliche Betrachtung der Thematik sei auf [RGRK04] verwiesen.

- die Rückgabe falscher Routing-Information. Auf diese Weise kann ein Angreifer die weitere Kommunikation eines Nutzers auf andere bösartige Knoten beschränken; das setzt voraus, dass der Angreifer von seinem Opfer überhaupt nach Routing-Informationen gefragt wird. Eine Diskussion dieses Angriffs und Maßnahmen zu seiner Verhinderung finden sich in [CDGR+02].

2.2 Grundlagen von Empfehlungssystemen

Empfehlungssysteme haben die Aufgabe,

- einem Nutzer Objekte zu empfehlen, die für ihn interessant sein könnten. Oft, aber nicht immer, sind dies Produkte, die der Nutzer erwerben kann.

- eine personalisierte Empfehlung von Objekten zu erstellen. Hierbei wird versucht, die Bewertung vorherzusagen, die der Nutzer dem Objekt geben würde, falls er es nach Benutzung selbst bewerten müsste. Diese Funktionalität wird benötigt, wenn ein Nutzer ein bestimmtes Objekt bereits in Erwägung zieht und wissen möchte, ob es für ihn geeignet wäre.

Im Rahmen der vorliegenden Arbeit ist die Eingabe des Empfehlungssystems eine Menge von *Bewertungen*, die Ausgabe eine *Empfehlung*.

Definition 3. Eine *Bewertung* ist die subjektive Einschätzung eines Objekts durch einen Nutzer. Die Ausgabe des Empfehlungssystems wird als *Empfehlung* bezeichnet.

Zur Erfüllung der genannten Aufgaben lassen sich folgende Ansätze unterscheiden [Pazz99]:

- *Inhaltsbasierte* Verfahren basieren darauf, dass ein Nutzer eine Beschreibung des gewünschten Objekts eingibt und das System diese mit Beschreibungen anderer Objekte vergleicht. Objekte mit ähnlichen Beschreibungen werden dem Nutzer als Empfehlungen präsentiert.
- *Demographische* Verfahren basieren auf demographischen Angaben. Ein Objekt wird dann empfohlen, wenn es von Nutzern des gleichen Geschlechts, ähnlichen Alters etc. benutzt wurde.
- Bei *Kollaborativem Filtern* werden Empfehlungen für einen Nutzer auf Basis der Bewertungen, die andere Nutzer über Objekte abgegeben haben, oder auf Basis des Nutzungsverhaltens anderer Nutzer erstellt.

Im Rahmen der vorliegenden Arbeit wird der Ansatz des Kollaborativen Filterns verfolgt. Dies liegt in den vielfältigen Vorteilen dieses Ansatzes begründet: So werden keine demographischen Angaben der Nutzer und keine (Meta-)Informationen über die zu empfehlenden Objekte benötigt. Auch kann Kollaboratives Filtern auf beliebige Objekte angewandt werden; es wird nicht gefordert, dass Objekte beispielsweise einem paarweisen Vergleich oder einer Beschreibung anhand gewisser Eigenschaften zugänglich sind. Verfahren, die den Ansatz des Kollaborativen Filterns umsetzen, bauen beispielsweise auf der Ähnlichkeit zwischen Nutzern auf; einem Nutzer werden dann diejenigen Objekte empfohlen, die von ähnlichen Nutzern eine positive Bewertung erhalten haben. Essentiell für die Funktionalität eines solchen Systems ist daher das verwendete Ähnlichkeitsmaß. In der Literatur wird beispielsweise der Korrelationskoeffizient nach Pearson herangezogen [KMMH+97, O'He99, SKKR00]; ebenfalls häufig verwendet wird der Cosinus des Winkels zwischen den – als Vektoren interpretierten – Nutzerprofilen [CoFa00, SKKR00]. Beide Maße stammen aus dem Gebiet des Information Retrieval, das mit dem der Empfehlungssysteme eng verwandt ist [BeCr92].

Grundlage eines Empfehlungssystems, das auf Kollaborativem Filtern beruht, ist eine Matrix ähnlich derjenigen, die in Abbildung 2.4 dargestellt ist: Eine Menge N von Nutzern gibt Bewertungen über eine Menge O von Objekten ab (statt Bewertungen sind auch reine Benutzungsvorgänge denkbar). Die Menge der Bewertungen zu einem Objekt o heißt nun B_o. Als Empfehlung ausgegeben wird

- für ein gegebenes Tupel (n, o) mit $n \in N$, $o \in O$: Die Vorhersage $b^*(n, o)$ der Bewertung des Objekts o durch den Nutzer n.
- für einen Nutzer n: Die Menge $E(n) \subseteq O$ diesem Nutzer empfohlener Objekte.

Von Empfehlungssystemen zu unterscheiden sind Reputationssysteme, die der Einschätzung der Vertrauenswürdigkeit von Nutzern dienen; dort wird für jeden *Nutzer n* eine Menge B_n Bewertungen von Bewertungen herangezogen.

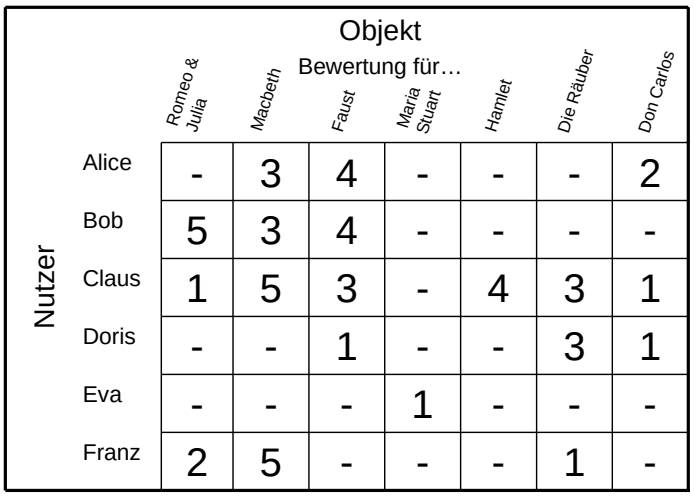

Abbildung 2.4: Nutzer-Objekt-Matrix

In einem verteilten Empfehlungssystem stellt sich nun die Frage, wie die Speicherung der Daten organisiert werden soll; in der Folge ergibt sich auch, welche Empfehlungsalgorithmen in Frage kommen.

Die dafür verwendbaren Ansätze lassen sich grundsätzlich in zwei Klassen einteilen.

Die erste Klasse beinhaltet nutzerbasierte Ansätze. Dies bedeutet, dass Bewertungen eines Nutzers gemeinsam gespeichert werden und dem Empfehlungsalgorithmus somit gemeinsam zur Verfügung gestellt werden können. In Abbildung 2.4 bedeutet dies, dass die Matrix zeilenweise gespeichert wird. Dies kann sowohl in einem strukturierten Peer-to-Peer-System als auch in einem unstrukturierten System geschehen. Der Ansatz eignet sich somit grundsätzlich auch für die Verwendung in Umgebungen mit unzuverlässigen Kommunikationsverbindungen, denn statt sich auf die Eigenschaften eines strukturierten Netzes zu verlassen, deren Aufrechterhaltung bei unzuverlässigen Kommunikationsverbindungen sehr aufwendig sein kann, genügt die Möglichkeit, Nachbarn aufzufinden - eine Aufgabe, die auch beim Wegfall bestimmter Verbindungen noch erfüllt werden kann. Die Empfehlungsqualität kann in diesem Fall jedoch leiden. Nutzerbasierte Ansätze sind bereits in der Literatur [PSWS05] vorgeschlagen worden. Die grundlegende Idee besteht darin, Algorithmen anzuwenden, die auch in zentralen Systemen Verwendung finden; hierbei werden dem Algorithmus aus der gesamten Nutzer-Objekt-Matrix (wie in Abbildung 2.4 dargestellt) nur einige Zeilen zur Verfügung gestellt – nämlich solche aus der (zu definierenden) Nachbarschaft des jeweiligen Knotens. Sogar für zentralisierte Empfehlungssysteme ist ähnliches vorgeschlagen worden, um den benötigten Rechenaufwand zu reduzieren [GRGP01]. In der konkreten Implementierung in [PSWS05] wird der Algorithmus des Amazon.com-Systems [LiSY03] eingesetzt. Für die Nachbarschaft eines Knotens aus Sicht des Empfeh-

lungssystems werden Peers ausgewählt, die dem suchenden Knoten möglichst ähnlich sind, die sogenannten *taste buddies*.

Die zweite Klasse beinhaltet objektbasierte Ansätze. Hier werden Bewertungen eines Objekts gemeinsam gespeichert, also die Spalten der Matrix aus Abbildung 2.4. Benötigt wird nun ein Weg, gezielt auf Daten über ein bestimmtes Objekt zugreifen zu können – die Betrachtung lediglich der Objekte, die beispielsweise auf benachbarten Knoten eines Nutzers verwaltet werden, reicht nicht aus. Die Möglichkeit, auf einem strukturierten Peer-to-Peer-System aufzubauen, schafft die notwendigen Voraussetzungen. In [MiKR04] werden verschiedene mögliche Architekturen vorgeschlagen, die für ein Empfehlungssystem benötigten Daten unter Verwendung einer DHT – konkret wird Chord [SMKK+01] herangezogen – zu speichern. Ähnlich zu einer dieser Architekturen wird in [WPLR06] ein System vorgestellt und evaluiert, das Musikempfehlungen auf Basis von Daten generiert, die in einem strukturierten Peer-to-Peer-System – auf das in [WPLR06] jedoch nicht eingegangen wird – abgelegt sind.

Problem dieses Ansatzes ist die mangelnde Skalierbarkeit in der Größe der Nutzerprofile – dementsprechend findet sich auch nur in [WPLR06] eine simulative Evaluierung, die dazu noch auf eine kleine Knotenzahl beschränkt ist. Da der Ansatz aber im Gegensatz zu nutzerbasierten Verfahren grundsätzlich ermöglicht, zu jedem im System bekannten Objekt Informationen – insbesondere Bewertungen – aufzufinden, soll in der weiteren Betrachtung ein grundlegendes Verfahren ähnlich dem in [WPLR06] verwendet werden. Es soll aber in seiner Skalierbarkeit, seinen Sicherheits- und Datenschutzeigenschaften erheblich verbessert werden. Ansätze zur Verbesserung der Skalierbarkeit sind in [Jäge06] beschrieben. Zur Erfüllung der gestellten Anforderungen reichen sie jedoch nicht aus.

2.2.1 Aufgaben und Anwendungsfälle

Um die gewünschte Funktionalität eines Empfehlungssystems zu erreichen, lassen sich zwei grundlegende Aufgabenstellungen identifizieren (vgl. dazu auch [HKTR04]).

- Zum einen sollen dem Nutzer zu einem Objekt, das er bereits kennt, Bewertungen angezeigt werden können. Neben der Möglichkeit, auf einzelne Bewertungen zugreifen zu können, soll auch eine Aggregation möglich sein. Diese Aggregation dient letztlich dazu, die Bewertung vorherzusagen, die der Nutzer einem Objekt selbst geben würde, und kann so eine Kaufentscheidung erleichtern. Einzelne Bewertungen sollten dabei als Dokumente in einem Peer-to-Peer-System gespeichert und eine Suche durch den Nutzer ermöglicht werden.

- Zum anderen soll der Nutzer aber auch auf Objekte aufmerksam gemacht werden, die er noch nicht kennt. Diese Aufgabe, die auch als „Finden relevanter Objekte“ bezeichnet wird, kann als Kernaufgabe eines Empfehlungssystems betrachtet werden.

2.2.2 Identifikation bewerteter Objekte

Um ein Objekt in einer DHT so speichern zu können, dass es effizient wieder aufgefunden werden kann, ist eine Identifikation dieses Objekts nötig. Die Existenz eines solchen Identifikators ist aber nicht selbstverständlich.

Für zahlreiche Produktklassen sind Identifikationsschemata jedoch bereits gebräuchlich; zu nennen sind beispielsweise

- ISBN (International Standard Book Number [Inte05]) für Bücher.
- ISRC (International Standard Recording Code [Inte01]) für Musikaufnahmen.
- GTIN (Global Trade Item Number) in verschiedenen Varianten für alle Arten von Handelswaren.
- DOI (Digital Object Identifier [Inte06]) für digitale Objekte und Objekte, zu denen Informationen im Internet abrufbar sind.

Trotz der Existenz dieser Schemata bleiben Probleme bei der Identifikation empfohlener Objekte:

- Die parallele Verwendung mehrerer Identifikationsschemata kann dazu führen, dass ein Objekt auf verschiedene IDs abgebildet wird. Diesem Problem kann begegnet werden, indem eine Rangfolge bevorzugter Identifikationsschemata definiert wird.
- Auch innerhalb eines Identifikationsschemas kann ein Objekt mehrere IDs haben.
- Verschiedene Versionen eines Objekts, die von den Nutzern als gleichwertig wahrgenommen werden, haben dennoch verschiedene IDs.

Begegnet werden kann diesen Problemen, indem in Dokumenten, die auf bestimmte Objekte Bezug nehmen, auf andere Identifikatoren des gleichen Objekts sowie auf verwandte Objekte hingewiesen wird. Das Durchsuchen des Netzes nach Stichwörtern [ReVa03] könnte für Nutzer weiterhin hilfreich sein, um ein bestimmtes Objekt zu identifizieren; dies wird jedoch im Rahmen dieser Arbeit nicht betrachtet. Ebenfalls außer Betracht bleibt die Bewertung von Personen bzw. anderen Nutzern des Systems; dies ist die Aufgabe von Reputationssytemen.

2.3 Ökonomischer Hintergrund: Eine Einführung

Empfehlungssysteme kommen in der Praxis meist dann zum Einsatz, wenn eine Entscheidung über den Kauf von Produkten an einem Markt gefällt werden muss; daher soll an

dieser Stelle ein kurzer Überblick der ökonomischen Sichtweise auf Empfehlungen gegeben werden.

Ein Markt ist ein Ort, an dem Angebot und Nachfrage zusammentreffen [BeKi99, S. 34]; er dient dem Austausch von Gütern, also Waren und Dienstleistungen. Nahezu alle materiellen Bedürfnisse können auf Märkten gedeckt werden. In einem einfachen Modell der Entscheidungsfindung eines am Markt agierenden Nachfragers hat dieser eine ihm bekannte Nutzenfunktion sowie beschränkte (finanzielle) Ressourcen. Am Markt gibt es eine Vielzahl an Anbietern, die ihm verschiedene Güter zu unterschiedlichen Preisen anbieten.

Der Nachfrager wird nun seine Ressourcen so einsetzen, dass der Nutzen maximiert wird, der ihm aus seinem Agieren am Markt entsteht. Diese Maximierung ist ihm jedoch nur möglich, wenn ihm hinreichende Informationen zur Verfügung stehen – was in der Regel nicht der Fall ist. Zu den benötigten Informationen zählen

- alle Eigenschaften der gehandelten Güter – dies betrifft auch ihre Qualität und die Benutzbarkeit für verschiedene Einsatzzwecke. Bereits Akerlof [Aker70] wies auf Informationsasymmetrien zwischen Käufer und Verkäufer hin. So ist es dem Käufer in vielen Fällen nicht möglich, hinreichende Informationen über das gewünschte Produkt zu erhalten.

- die Anbieter der Güter und die von diesen Anbietern verlangten Preise inklusive entstehender Transaktionskosten.

- die eigenen Präferenzen, bezogen auf die angebotenen Güter. Anders ausgedrückt: Der Nachfrager muss zu jedem Produkt wissen, wieviel es zu seinem eigenen Nutzen beitragen würde, um eine Entscheidung für oder gegen den Kauf dieses Produkts treffen zu können.

Insgesamt ist also, neben dem eigentlichen Treffen der Kaufentscheidungen, auch die Bereitstellung der dafür notwendigen Informationen problematisch. Auch gilt zu beachten, dass bereits die Suche nach einer optimalen Lösung Ressourcen beansprucht (Suchkosten), so dass das Finden einer nahezu optimalen Lösung sich im Allgemeinen nicht lohnt.

Es liegt nun nahe, die Optimierungsaufgabe eines Nachfragers auf elektronischem Weg zu unterstützen. Bereits die Verbreitung von Online-Shops, die mit Hilfe von Suchmaschinen aufgefunden werden können, war ein erster Schritt in diese Richtung. Je mehr Produktinformationen über Rechnernetze abrufbar sind, desto leichter wird – zumindest bei geeigneter Darstellung dieser Produktinformationen – die Informationsgewinnung. Insbesondere die Information über Anbieter und Preise wird erleichtert; Suchkosten werden durch den Electronic Commerce somit gesenkt [Bako97]. In einem nächsten Schritt kann die Informationsgewinnung oder letztlich gar der Kauf beispielsweise an Softwareagenten delegiert werden, was die Transaktionskosten weiter senken kann.

Empfehlungen können nun als weitere Unterstützung des Such- und Entscheidungsprozesses verstanden werden. Die Bedeutung von Bewertungen anderer Nutzer für den Entscheidungsfindungsprozess wird auch in der ökonomischen Literatur betont. In [AvRZ99] wird aus diesem Grund beispielsweise sogar das Schaffen eines Marktes für die Bewertungen selbst vorgeschlagen. In der vorliegenden Arbeit wird zwar nicht von einem solchen Markt ausgegangen; auch ohne einen solchen können aber durchaus zahlreiche Vorteile durch ein Empfehlungssystem erreicht werden:

- Durch Empfehlungen kann der Nachfrager von der Existenz eines Produkts erfahren.
- Auch können Empfehlungen Informationen über die Eigenschaften eines Produkts beinhalten.
- Schließlich kann ein Empfehlungssystem sogar helfen, die eigene, zu maximierende Nutzenfunktion zu bestimmen. So könnte ein Konsument beispielsweise aufgrund seiner Interessen dahingehend beraten werden, ein bestimmtes Produkt zu kaufen – auch, wenn ihm dieses Produkt vorher gar nicht bekannt war.

2.4 Fazit

Im vorliegenden Kapitel wurden die Grundlagen aufgezeigt, die für den Entwurf eines dezentralen, selbstorganisierenden Empfehlungssystems benötigt werden. Aus Sicht der Telematik sind dies insbesondere die Schichtenmodelle der Telekommunikation, der Einsatz kryptographischer Verfahren, um sichere Kommunikation zu erzielen, sowie die Erbringung von Diensten durch Peer-to-Peer-Systeme. Daneben wurden auch die Grundlagen von Empfehlungssystemen dargestellt sowie einige grundlegende Aspekte von Empfehlungen aus Sicht der Ökonomie beleuchtet.

Kapitel 3

Datenschutz

In Kapitel 2 wurden die Grundlagen eines dezentralen und selbstorganisierenden Empfehlungssystems skizziert. Doch abgesehen von der – rein technisch betrachteten – Funktionalität des Systems müssen an eine Anwendung der Telekommunikation sowohl aus Nutzersicht als auch aus Sicht des Rechtssystems weitere Anforderungen gestellt werden. Diese werden im vorliegenden Kapitel erörtert.

3.1 Motivation

Es versteht sich von selbst, dass die Funktionalität eines Empfehlungssystems die Verarbeitung von Daten erfordert. Zum einen sind dies Daten, die technisch bedingt in einem Kommunikationssystem grundsätzlich anfallen, wie z.B. die Adressen beteiligter Endsysteme. Zum anderen werden Daten verarbeitet, die für die eigentliche Funktionalität des Empfehlungssystems von Belang sind. Auf den ersten Blick mag diese Datenverarbeitung unproblematisch erscheinen, da der Sinn des Systems in der Bewertung von Produkten, nicht von Nutzern besteht. Jedoch kann die Bewertung von Nutzern sich durchaus als hilfreich herausstellen, wenn es gilt, eine Einschätzung über die Qualität von Empfehlungen zu gewinnen. Zudem ist die Information, wie ein Nutzer ein Produkt bewertet hat, nicht allein eine Information über das Produkt, sondern auch über den Nutzer. Mag dies bei einer einzelnen Bewertung noch unkritisch sein, so kann die Gesamtheit aller Bewertungen, die ein Nutzer abgegeben oder angefordert hat, doch bereits Rückschlüsse auf diesen Nutzer zulassen, die von diesem möglicherweise nicht gewünscht sind. So können Empfehlungen im Bereich von Nahrungsmitteln oder pharmazeutischen Produkten Hinweise auf Vorlieben oder gar Erkrankungen des Nutzers geben. Aber auch scheinbar banale Daten wie Kalenderdatum und Uhrzeit, zu der eine Empfehlung erstellt wurde, lassen auf Gewohnheiten des Empfehlenden schließen.

Ein derartige Möglichkeit zur Verletzung der Privatsphäre eines Nutzers sollte in einem Empfehlungssystem minimiert werden. Es gilt, eine möglichst gute Funktionalität des Systems bei gleichzeitigem Schutz seiner Nutzer vor ungewollter oder nicht notwendiger Preisgabe ihrer Daten zu erreichen.

Aus diesem Grund wird im Folgenden analysiert, wie die Anforderungen an den Datenschutz in einem Empfehlungssystem aus Nutzersicht aussehen können. Anschließend werden Möglichkeiten diskutiert, diese Anforderungen mit Hilfe des Rechts, aber auch mit technischen Mitteln zu erreichen: Es ist davon auszugehen, dass weder technische noch juristische Maßnahmen jeweils alleine zur Sicherstellung eines ausreichenden Datenschutzniveaus in der Lage sind. Zwar können technische Maßnahmen den sparsamen Umgang mit personenbezogenen Daten fördern, aber nur schwer den Missbrauch einmal erhobener Daten verhindern. Das Recht wiederum kann solche Regelungen aufstellen, bedarf jedoch der Verwirklichung in einer konkreten technischen Umsetzung.

3.1.1 Anforderungen an den Datenschutz

Datenschutz dient dazu, natürliche Personen in Bezug auf den Umgang mit ihren personenbezogenen Daten zu schützen. Ausgangspunkt der Betrachtung sollen deshalb weder die Rechtslage noch technische Schutzmaßnahmen sein. Vielmehr werden zunächst die Anforderungen der Nutzer an den Datenschutz betrachtet. Empirische Untersuchungen zu Nutzeranforderungen an den Datenschutz in Empfehlungssystemen liegen noch nicht vor; jedoch können allgemeine Untersuchungen zu Datenschutzanforderungen herangezogen werden, um daraus im nächsten Schritt eine Spezialisierung auf Empfehlungssysteme abzuleiten. Erst im Anschluss wird dann betrachtet, welche Anforderungen an den Datenschutz aus Sicht des Rechts zu stellen sind und wie diese einerseits wieder durch gesetzgeberische Maßnahmen, andererseits durch technische Verfahren erfüllt werden können.

3.1.1.1 Allgemeine Nutzeranforderungen

Verschiedene Studien haben gezeigt, dass Datenschutz von der Bevölkerung im allgemeinen und von Nutzern des Internets – mit Schwerpunkt auf dem World Wide Web, was auf dessen Beliebtheit zurückzuführen ist – im speziellen als wichtig angesehen wird. In einer Studie von Harris Interactive [Harr01, Q705] antworteten beispielsweise 79,7% der Befragten (US-Bürger über 18 Jahren), es sei ihnen „sehr wichtig“, die Kontrolle darüber zu haben, welche Informationen über sie gesammelt werden. In Deutschland wünschen sich nach einer Umfrage des Freizeit-Forschungsinstituts 53% der Befragten, dass dem Datenschutz künftig mehr Bedeutung zukommt [Opas01, Teil VI]. Bemerkenswert erscheint, dass die Bedenken der Internetnutzer in bezug auf Datenschutz mit wachsender Erfahrung zunehmen [MiFe01, S. 38]. Dies deutet darauf hin, dass nicht etwa Berührungsängste oder Unkenntnis – beispielsweise bezogen auf mögliche Schutzmaßnahmen, Verschlüsselungsmechanismen und Datenschutzprotokolle – Ursache dieser Bedenken sind. Vorausgesetzt wird bei dieser Folgerung allerdings, dass mit der Erfahrung eines Internetnutzers auch dessen technisches Wissen zunimmt.

Doch wann sehen Nutzer die Weitergabe von Daten wirklich als problematisch an? Ausgehend von dieser Fragestellung werden in der Literatur verschiedene Einflussfaktoren auf den Datenschutz diskutiert. Für die Zwecke dieser Arbeit wird auf dieser Grundlage ein Modell erstellt, in dem die von Nutzern verfolgten Schutzziele in fünf Kategorien (Ebenen) eingeteilt werden. Entsprechend der Ansätze, diese Ziele zu unterstützen, findet sich ein Teil dieser Kategorien in Gesetzgebung und Rechtsprechung wieder, ein anderer Teil in technischen Ansätzen zum Datenschutz. Im Einzelnen handelt es sich um

- die *Inhaltsebene*, in der betrachtet wird, welche Arten von Daten der Nutzer preiszugeben bereit ist,
- die *Empfängerebene*, deren Gegenstand die Auswahl von potentiellen Empfängern der Daten ist,
- die *Identitätsebene*, in der es um den Schutz der Identität des Nutzers geht,
- die *Zweckebene*, in der der Verwendungszweck der Daten betrachtet wird,
- die *Kontrollebene*, die sich auf die Kontrolle des Nutzers über erhobene Daten bezieht.

Diese Ebenen werden nun näher betrachtet. Dabei gilt jedoch zu beachten, dass die Bereitschaft eines Nutzers zur Preisgabe persönlicher Daten nicht lediglich durch eine Ebene bestimmt wird; stattdessen spielen jeweils mehrere oder alle Ebenen zusammen. Auch spielen individuelle Neigungen der Nutzer eine Rolle; diese entziehen sich jedoch dem Einfluss des Systemgestalters und werden deshalb an dieser Stelle nicht näher betrachtet.

Inhaltsebene Hier gilt es, über Art und Umfang preiszugebender Daten zu entscheiden. Offensichtlich werden nicht alle Daten als gleich sensibel empfunden. So gaben in einer Umfrage (s. [AcCR99, S. 3]; basierend auf [CrRA99, Anhang, Frage 12]), 82% der (überwiegend amerikanischen) Befragten an, die Preisgabe ihrer bevorzugten Fernsehsendung gegenüber einer Website üblicherweise oder immer als unproblematisch zu empfinden; bei der Postanschrift galt dies nur für 44%, bei medizinischen Informationen für 18% und bei der – in den USA besonders sensiblen – Sozialversicherungsnummer nur noch für 1%.

Empfängerebene Die Bereitschaft, personenbezogene Daten preiszugeben, hängt auch vom Gegenüber ab. Eine mögliche Ursache hierfür ist die wahrgenommene Zuverlässigkeit des Empfängers im Umgang mit den Daten. So gaben in der bereits zitierten Umfrage des Freizeit-Forschungsinstituts [Opas01, Teil VI] 52% der Befragten an, Banken für „absolut zuverlässig“ im Umgang mit privaten Daten zu halten, während dies für Internetanbieter[1] und Versandhändler nur noch 10% annahmen.

[1] Dieser unscharfe Begriff wurde laut [Opas01, Teil VI] in der Umfrage verwendet.

Innerhalb der Empfängerebene gilt zu differenzieren, wie die Nutzer den Empfängerkreis einschränken wollen:

- Sollen Daten nur gegenüber einem genau bestimmten Empfängerkreis offenbart und ihre nicht autorisierte Weitergabe verhindert werden?
- Soll lediglich ganz allgemein verhindert werden, dass „zu viele“ Empfänger Daten über den Betroffenen erhalten?
- Will der Betroffene nur verhindern, dass ein einzelner Empfänger zu viele Daten über ihn kennt? Dies ist durchaus vorstellbar, da oft erst eine hinreichend große Datenmenge das Erschließen Informationen über eine Person erlaubt.

Inwieweit diese Unterscheidung in der Praxis eine Rolle spielt, ist bislang empirisch nicht untersucht; diese Fragestellung könnte für zukünftige Forschungsarbeiten aber von Relevanz sein.

Neben der Frage nach dem Empfängerkreis spielt auch die Datensicherheit in der Empfängerebene eine Rolle, wenn die unbeabsichtigte Preisgabe von Informationen berücksichtigt werden soll (vgl. dazu [PaJø04, S. 7]).

Identitätsebene In der Identitätsebene wird betrachtet, ob Daten einem bestimmten Nutzer zugeordnet werden können – umgekehrt stellt sich für den Nutzer die Frage, ob er seine Identität preisgibt. Somit entscheidet sich, ob preisgegebene Daten Personenbezug haben. Doch kann keine binäre Unterscheidung (Bekanntgabe oder Geheimhaltung der Identität) getroffen werden. So wird beispielsweise in [ReRu98, S. 69] eine Skala eingeführt, die sich an der Wahrscheinlichkeit orientiert, mit der Aussagen über die Identität einer Person (dort spezifisch als Teilnehmer eines Kommunikationsvorgangs) zu treffen sind. Neben die Frage, *ob* die Identität preisgegeben wird, kann die Frage treten, mit *welcher* Identität der Nutzer auftritt. So kann eine Person unter einer Vielzahl verschiedener Pseudonyme agieren. Die Verwendung eines Pseudonyms bedeutet jedoch nicht notwendigerweise die Lösung aller Datenschutzprobleme. So können Pseudonyme langlebig und für ihren Träger sehr wertvoll sein – man betrachte nur die Pseudonyme, unter denen ein Verkäufer bei eBay auftritt und die im dort verwendeten Reputationssystem [ReZe02] zugrunde gelegt werden. Auch besteht die Gefahr, dass verschiedene Identitäten einer Person zusammengeführt werden können.

Als ein Beispiel, wie dieses Problem von Nutzern des Internet gesehen wird, kann der Umgang mit Cookies dienen: Diese können dazu eingesetzt werden, die sonst oft nur kurzlebigen Identitäten der Nutzer einer Website durch Vergabe eines eindeutigen, auf Client-Seite gespeicherten Identifikators zu einer langlebigen Identität zu verknüpfen. Eine Umfrage [AcCR99, S. 5], [CrRA99, Anhang, Frage 18] ergab, dass 52% der Web-Nutzer die Verwendung von Cookies als problematisch ansehen; ein Ergebnis, dessen Aussagekraft jedoch

durch Unkenntnis und falsche Vorstellungen der Nutzer, die von den Autoren der Studie konstatiert werden, eingeschränkt wird. Die Frage nach einem persistenten Identifikator wurde in der gleichen Umfrage ohne Verwendung des Begriffs „Cookie“ erneut gestellt; nun waren zunächst 78% mit einem solchen Identifikator einverstanden. Es zeigt sich jedoch die Interdependenz mit Zweck- und Empfängerebene: Die Verwendung des Identifikators für personalisierte Werbung fand noch die Zustimmung von 60% der Befragten. Sollte die langlebige Identität zum Zweck personalisierter Werbung gegenüber vielen Websites verwendet werden, stimmten nur noch 44% zu.

Auch zeigt sich – wenig überraschend –, dass Nutzer eher bereit sind, Informationen preiszugeben, wenn diese nicht mit ihrem Namen verknüpft sind: So wurden amerikanische Internetnutzer nach ihrer Bereitschaft gefragt, für personalisierte Dienste ihre Postleitzahl anzugeben und einige Fragen zu ihren Interessen zu beantworten. 83,5% waren dazu „definitiv“ oder „wahrscheinlich“ bereit. Diese Zahl sank auf 49,5%, wenn zusätzlich der Name angegeben werden sollte. [CrRA99, Anhang, Fragen 3 und 4].

Die Wahrnehmung der Bedeutung der Identitätsebene deckt sich mit den Bemühungen der Forschung, Datenschutz in Rechnernetzen zu gewährleisten – diese setzen zu einem großen Teil auf der Identitätsebene an (vgl. Abschnitt 3.3). Auch in der vorliegenden Arbeit liegt der Schwerpunkt der datenschutzgerechten Umsetzung eines Empfehlungssystems auf der Identitätsebene.

Zweckebene Die Zweckebene betrifft den Verwendungszweck der Daten, zu dem diese offenbart werden. Wünscht der Nutzer die Beschränkung der Verwendung der Daten auf einen bestimmten Zweck? Gibt er sie nur preis, wenn er davon ausgehen kann, dass es bei der Verwendung zu diesem Zweck bleibt? Wie sonst soll er sich vor missbräuchlicher Verwendung schützen?

Es zeigt sich, dass Internetnutzer die Zweckbindung sehr ernst nehmen. 97,3% der befragten Internetnutzer [CrRA99, Anhang, Frage 20][2] bezeichneten den beabsichtigten Verwendungszweck als wichtiges oder sehr wichtiges Kriterium bei der Bewertung einer Privacy Policy. Dies deckt sich mit dem Ergebnis einer anderen Umfrage [KrLG02, S. 115,Q 211]: Demnach betrachten 53% die Möglichkeit, ihre Daten könnten außerhalb der Transaktion, für die sie gedacht waren, verwendet werden, mit „großer Sorge“ (im Original: „a major concern“) und weitere 42% mit „Sorge“ („a minor concern“). Lediglich 5% sahen hier überhaupt kein Problem.

Kontrollebene Die letzte betrachtete Ebene ist die Kontrollebene. Wie beeinflusst einen Nutzer die Kontrolle, die er über preisgegebene Daten hat? Zum einen ist dies eine Frage der Transparenz: Kontrolle kann nur stattfinden, wenn der Nutzer weiß, welche Daten er of-

[2] In neueren Untersuchungen wurde diese Fragestellung allerdings nicht erneut aufgegriffen.

fenbart, wo diese gespeichert und wie sie verarbeitet werden. So wünschten sich beispielsweise 61,4% der befragten amerikanischen Internetnutzer die Möglichkeit, Formulare auf Websites *auf Anforderung* automatisch mit dem Browser bereits bekannten Informationen auszufüllen – ein Wunsch, der mit sinkender Transparenz der Implementierung zurückging: Ein Ausfüllen ohne Anforderung, bei dem die Formulardaten erst durch Nutzerinteraktion versendet werden, wurde nur noch von 50,9% gewünscht; ein automatisches Versenden ohne Eingrifssmöglichkeit, aber mit Information des Nutzers wünschten noch 13,9%, und ein unbemerktes Versenden von Informationen wurde nur noch von 5,5% der Nutzer begrüßt.

Zum anderen bedeutet Kontrolle aber auch, Einfluss auf Daten nehmen zu können, auch, wenn diese bereits offenbart wurden – beispielsweise, um fehlerhafte Informationen zu korrigieren oder auch die Löschung von Daten zu veranlassen. Dazu gehört auch der Wunsch nach Sanktionen, wenn mit Daten nicht in der erwarteten Form umgegangen wird.

In einer Studie [Fox00, S. 11] wünschten beispielsweise 94% der Befragten solche Sanktionen, falls der Betreiber einer Website personenbezogene Daten in einer anderen Weise als angegeben verarbeitet – 11% wünschten Gefängnisstrafen, 26% ein Schließen der betreffenden Website und 27% sonstige Strafen gegen den Betreiber. Immerhin 30% wünschten jedoch lediglich einen Reputationsmechanismus und fordern die Veröffentlichung einer Liste mit betrügerischen Websites. Weitere 2% wünschten keine Bestrafung; weitere Antwortmöglichkeiten waren nicht vorgesehen.

Die Bedeutung der Kontrollebene wird im Zusammenspiel mit den anderen Ebenen deutlich – beispielsweise, wenn die Kontrolle über den Verwendungszweck von Daten betrachtet wird: Ein Nutzer kann bereit sein, personenbezogene Daten für einen bestimmten Zweck (wie z.B. die Teilnahme an einem Gewinnspiel) preiszugeben. Doch die Gewissheit, dass die Daten auch tatsächlich nur für diesen Zweck verwendet werden, erhält der Nutzer nur durch den Einsatz von Kontrollmechanismen.

Die entwickelte Einordnung der Datenschutzanforderungen in fünf Ebenen wird für den Entwurf eines datenschutzgerechten Empfehlungssystems erneut aufgegriffen werden.

3.2 Grundlagen des Datenschutzes im deutschen Recht

Die bestehenden Anforderungen der Bevölkerung an den Datenschutz sind kein neues Phänomen. Schon 1890 diskutierten Warren und Brandeis [WaBr90] – bezogen auf amerikanisches Recht – ein „right to privacy“, also ein Recht auf Privatsphäre. Die Autoren kommen in ihren Überlegungen dem heutigen Konzept des Datenschutzes bereits recht nahe.

Nachdem in den 1950er Jahren der Einsatz elektronischer Datenverarbeitung begonnen hatte, erkannten die deutschen Gesetzgeber, beginnend 1969, die Datenschutzproblematik;

in Hessen wurde das erste Landesdatenschutzgesetz[3] erlassen, und nach langen Debatten trat das Bundesdatenschutzgesetz 1978 in Kraft (Abel in [Roßn03, Abschnitt 2.7, Rn. 13–17]).

Wesentlich geprägt wurde das deutsche Datenschutzrecht durch das Volkszählungsurteil des Bundesverfassungsgerichts [Bund83]. In diesem Urteil wurden die verfassungsrechtlichen Anforderungen an die Verarbeitung personenbezogener Daten formuliert. Insbesondere leitete das Gericht das *informationelle Selbstbestimmungsrecht* – also die „Befugnis des Einzelnen, grundsätzlich selbst zu entscheiden, wann und innerhalb welcher Grenzen persönliche Lebenssachverhalte offenbart werden“ [Bund83] – aus Artikel 2 Abs. 1 in Verbindung mit Artikel 1 Abs. 1 Grundgesetz (GG) ab. Mit dieser Formulierung knüpft das Bundesverfassungsgericht an Westins Definition des Begriffs „Privacy“ aus dem Jahre 1967 an („Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others“ [West67, S. 7]). Das informationelle Selbstbestimmungsrecht ist jedoch nicht schrankenlos gewährleistet; dies ergibt sich schon aus der Einbindung des Einzelnen in die soziale Gemeinschaft [Bund83]. Einschränkungen, die der deutsche Gesetzgeber vorgenommen hat, dienen beispielsweise der Strafverfolgung (wie die Rasterfahndung §§ 98a, 98b der Strafprozessordnung StPO).

Weitere Meilensteine auf dem Weg zum heutigen Datenschutzrecht waren die Novellierungen des Bundesdatenschutzgesetzes in den Jahren 1990 und – zur Umsetzung der europäischen Datenschutz-Richtlinie – 2001 (Abel in [Roßn03, Abschnitt 2.7, Rn. 44,52]). Bereichsspezifische Regelungen aufgrund der spezifischen Herausfordeurngen einzelner Regelungsgebiete – wie das 1997 verabschiedete Teledienstedatenschutzgesetz (TDDSG), das 2007 in das neu geschaffene Telemediengesetz integriert wurde – ergänzen das Bundesdatenschutzgesetz.

3.2.1 Allgemeines

Kern des deutschen Datenschutzrechts ist das bereits erwähnte Bundesdatenschutzgesetz. Es ist als sogenanntes Auffanggesetz konzipiert. Wenn und soweit bereichsspezifische Regelungen zum Datenschutz bestehen, so haben diese Vorrang (§ 1 Abs. 3 S. 1 BDSG); andernfalls ist das BDSG anzuwenden. Beispiele solcher bereichsspezifischen Regelungen sind die §§ 91–107 TKG (Telekommunikationsgesetz), das Teledienstedatenschutzgesetz (TDDSG) bzw. die aus diesem in das Telemediengesetz (TMG) übernommenen Regelungen. Sieht man von Sonder-, Schluss- und Übergangsvorschriften ab, so ist das Gesetz in drei Abschnitte gegliedert; sie enthalten

- allgemeine und gemeinsame Bestimmungen (also solche, die für öffentliche und für nicht-öffentliche Stellen gelten)
- Regelungen über die Datenverarbeitung der öffentlichen Stellen und

[3] Landesdatenschutzgesetze spielen für das Themengebiet der vorliegenden Arbeit allerdings keine Rolle.

- Regelungen über die Datenverarbeitung nicht-öffentlicher Stellen und öffentlich-rechtlicher Wettbewerbsunternehmen.

Grundlegende Regelungen des Gesetzes sollen an dieser Stelle nur kurz angerissen werden; ausführlichere Darstellungen finden sich in [WoGe05] und [GoKl03].

Normadressat (also durch das Gesetz angesprochene Personen) sind – neben den hier nicht betrachteten öffentlichen Stellen – alle nicht-öffentlichen Stellen, 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“ (§ 1 Abs. 2 Nr. 3 BDSG), sofern dies nicht ausschließlich im Rahmen einer persönlichen oder familiären Tätigkeit geschieht. Nicht-öffentliche Stellen sind dabei grundsätzlich alle „natürliche[n] und juristische[n] Personen, Gesellschaften und andere Personenvereinigungen des privaten Rechts“. Der sachliche Anwendungsbereich erstreckt sich auf die Verarbeitung, Nutzung und Erhebung *personenbezogener* Daten (§ 1 Abs. 2 BDSG), also „Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person“ (§ 3 Abs. 1 BDSG).

Eine grundlegende Regelung des BDSG, die sowohl öffentliche als auch nicht-öffentliche Stellen betrifft, ist das Verbot mit Erlaubnisvorbehalt des § 4 Abs. 1. Demnach dürfen personenbezogene Daten nur erhoben, verarbeitet oder genutzt werden, wenn ein Gesetz dies erlaubt oder der Betroffene eingewilligt hat; die Einwilligung bedarf dabei grundsätzlich der Schriftform (§ 4a Abs. 1 S. 3 BDSG). Weiterhin wird der Grundsatz der Erhebung beim Betroffenen aufgestellt (§ 4 Abs. 2 BDSG). Dies bedeutet, dass Daten (bis auf wenige Ausnahmen) direkt beim Betroffenen und nicht bei Dritten erhoben werden dürfen; dies dient der Transparenz der Datenerhebung.

Für den nicht-öffentlichen Bereich ist § 28 die zentrale Erlaubnisnorm. Die Erhebung, Speicherung, Veränderung oder Übermittlung wird unter anderem dann für zulässig erklärt, wenn dies der Zweckbestimmung eines Vertragsverhältnisses mit dem Betroffenen dient (Abs. 1 S. 1 Nr. 1), es zur Wahrung berechtigter Interessen der verantwortlichen Stelle erforderlich ist (Abs. 1 S. 1 Nr. 2) oder die Daten allgemein zugänglich sind (Abs. 1 S. 1 Nr. 3). In den letzten beiden Fällen ist dabei eine Abwägung mit den schutzwürdigen Interessen des Betroffenen erforderlich. Stets ist zu beachten, dass bereits bei der Erhebung der Zweck der Verarbeitung oder Nutzung konkret festzulegen ist (Abs. 1 S. 2).

Weitere Erlaubnistatbestände des Abschnitts über nicht-öffentliche Stellen betreffen unter anderem Werbung (vgl. Abschnitt 3.7), Markt- und Meinungsforschung und geschäftsmäßige Datenerhebung zum Zweck der Übermittlung. Auf die letztgenannten soll an dieser Stelle jedoch nicht eingegangen werden.

Zu erwähnen ist hingegen der für die Transparenz der Datenspeicherung zentrale Auskunftsanspruch des Betroffenen. Nach § 34 Abs. 1 BDSG kann er grundsätzlich Auskunft über die zu seiner Person gespeicherten Daten, die Empfänger, an die diese Daten weiterge-

geben wurden und den Zweck der Speicherung verlangen. Die Auskunft ist dabei grundsätzlich schriftlich zu erteilen (§ 34 Abs. 3).

3.2.2 Datenschutz in der Telekommunikation

Bereits das Bundesdatenschutzgesetz ist eine Reaktion auf die Möglichkeiten und Risiken, die durch die Verarbeitung personenbezogener Daten in Rechnersystemen entstehen. Doch ergeben sich durch die Vernetzung solcher Systeme neue Risiken – insbesondere, seit die Nutzung vernetzter Rechnersysteme sich verbreitet hat und auch Privatnutzern in großem Umfang zugänglich ist. Aus diesem Grund wurden bereichsspezifische Regelungen nötig, die der Gesetzgeber schuf

- durch Verabschiedung des Teledienstedatenschutzgesetzes (TDDSG) 1997,
- durch Aufnahme von – zum TDDSG fast wortgleichen – Datenschutzregelungen in den ebenfalls 1997 verabschiedeten Mediendienste-Staatsvertrag[4],
- durch Regelungen zum Telekommunikationsdatenschutz, die – nachdem sie sich zuvor in der Telekommunikations-Datenschutzverordnung (TDSV) gefunden hatten – im Jahr 2004 in das Telekommunikationsgesetz (TKG) aufgenommen wurden.

Soweit in diesen Gesetzeswerken keine Regelungen getroffen sind, gelten die Regelungen des BDSG. Im folgenden Abschnitt wird die Abgrenzung der genannten Regelungswerke diskutiert.

3.2.2.1 Ein Schichtenmodell des Datenschutzes in der Telekommunikation

Rechnernetze werden üblicherweise gemäß einem Schichtenmodell strukturiert (vgl. Abschnitt 2.1.1). Auf diesem Ansatz aufbauend, wird auch in der juristischen Literatur die Abgrenzung der verschiedenen Regelungsmaterien des Datenschutzes in der Telekommunikation mit Hilfe eines Schichtenmodells diskutiert (siehe nur [Schl04], [GoMü00, S. 105 ff.]). Dieses Schichtenmodell ist jedoch, wie unten im Einzelnen dargestellt wird, nicht deckungsgleich mit dem ISO/OSI-Schichtenmodell oder dem TCP/IP-Modell (anderer Auffassung Raabe in [Raab03, S. 136]). Vielmehr werden alle Aspekte der reinen Übertragung von Nachrichten in einer Schicht zusammengefasst. Eine graphische Darstellung des Datenschutz-Schichtenmodells im Vergleich zum TCP/IP-Schichtenmodell findet sich in Abbildung 3.1. Sie verdeutlicht, dass sich die Grenze zwischen den Schichten des Datenschutz-Schichtenmodells nicht trennscharf entlang der Grenzen des TCP/IP-Schichtenmodells[5] ziehen lässt. Die Interaktionsschicht des Datenschutzmodells beginnt jedenfalls oberhalb der Transportschicht des TCP/IP-Modells. Ob die Anwendungsschicht

[4] Die Regelungen des TDDSG und die Datenschutzregelungen des MdStV wurden 2007 in das TMG überführt.
[5] Und somit auch nicht des ISO/OSI-Schichtenmodells.

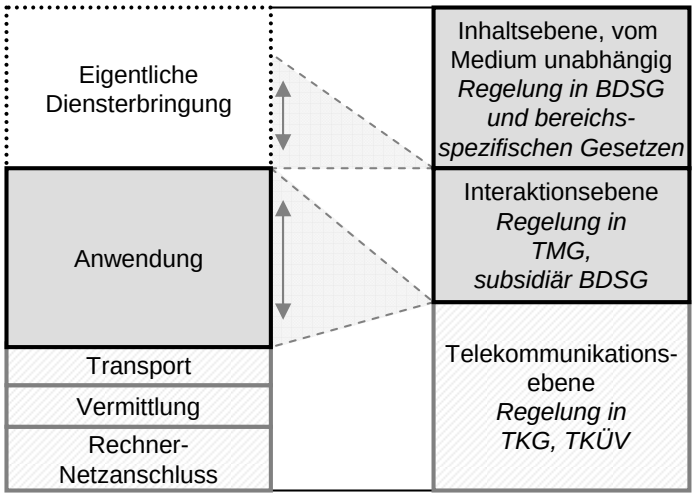

Abbildung 3.1: Telekommunikationsdatenschutz im Schichtenmodell: Links das TCP/IP-Modell, rechts das Datenschutz-Modell. Nutzerinteraktion und Inhalte der Kommunikation (gestrichelt dargestellt) sind im eigentlichen TCP/IP-Modell nicht enthalten.

oder Teile davon ebenfalls noch zur Interaktionsebene zu rechnen sind, lässt sich nur im Einzelfall entscheiden. In aller Regel wird ein Dienst jedoch nicht nur auf der Anwendungsschicht des TCP/IP-Modells erbracht, denn dieses Modell befasst sich lediglich mit Kommunikationsaspekten. In Abbildung 3.1 ist deshalb oberhalb der Anwendunschicht eine weitere Schicht eingezeichnet, die rechtlich je nach Einzelfall zumindest teilweise ebenfalls noch der Interaktionsebene zuzurechnen ist. Oberhalb der Interaktionsebene liegt schließlich die Inhaltsebene.

Die einzelnen Schichten lassen sich dabei wie folgt abgrenzen:

- *Schicht 1: Telekommunikation* Die unterste Schicht des Schichtenmodells umfasst alle Aspekte der Telekommunikation, die definiert ist als „der technische Vorgang des Aussendens, Übermittelns und Empfangens von Signalen mittels Telekommunikationsanlagen“ (§ 3 Nr. 22 TKG). Diese Definition ist zwar sehr umfassend, doch aufgrund des verwendeten Begriffs eines „technischen Vorgangs“ geht die herrschende Meinung von einer funktionalen Abgrenzung aus: Das TKG umfasst demnach nur den Transport von Nachrichten – also eine rein technische Dienstleistung –, nicht jedoch deren inhaltliche Verarbeitung ([Schm00, S. 71], Neubauer in [MoDr05, Teil D, Rz. 6]). Der Telekommunikationsbegriff des TKG lässt sich nicht trennscharf an Schichtengrenzen des TCP/IP-Referenzmodells abgrenzen (anderer Auffassung Raabe in [Raab03, S. 136], der Dienste der Anwendungsschicht als Teledienste einordnet. Zum gleichen Ergebnis kommen Helmke et al. [HeMN98, Rz. 44], allerdings vom ISO/OSI-Modell ausgehend. Teils wird die Grenze auch zwischen Schicht 5 und 6

des ISO/OSI-Modells gezogen, so in [Hoer98, S. 4] bezüglich des Begriffs der Telekommunikationsanlagen[6]). So ist der Dienst E-Mail mit dem SMTP-Protokoll in die Anwendungsschicht des Referenzmodells einzuordnen, doch besteht hier kein inhaltlicher Bezug; der reine Transport einer E-Mail unterfällt also dem Begriff der Telekommunikation (vgl. Spindler in [SpSG04, Rn. 50 zu § 2 TDG]). Die Grenze zwischen Transport- und Anwendungsschicht des TCP/IP-Referenzmodells und zwischen Telekommunikations- und Inhaltsebene fallen vielmehr nur dann zusammen, wenn auf der Anwendungsschicht keine Weiterleitung von Nachrichten mehr erfolgt. Am ehesten ist dies bei Protokollen der Anwendungsschicht gegeben, bei denen ein Client direkt mit einem Server kommuniziert. Die dabei auf Anwendungsschicht erbrachte Dienstleistung kann weit über den reinen Transport von Nachrichten hinausgehen[7]. Anders stellt sich die Situation beispielsweise bei der Overlay-Schicht des SESAM-Basissystems [CDHS+05] dar, die zwar zumindest teilweise der Anwendungsschicht des TCP/IP-Modells zugerechnet werden kann, aber aus rechtlicher Sicht einen reinen Transportdienst darstellt und somit ausschließlich als Telekommunikation eingeordnet werden kann. In Abbildung 3.1 stimmen die Schichtengrenzen des TCP/IP-Referenzmodells aus diesen Gründen nicht mit denen des Datenschutz-Schichtenmodells überein.

Der Datenschutz auf der Telekommunikationsebene ist in Teil 7 des TKG geregelt. Abschnitt 1 regelt das Fernmeldegeheimnis, dem „der Inhalt der Telekommunikation und ihre näheren Umstände, insbesondere die Tatsache, ob jemand an einem Telekommunikationsvorgang beteiligt ist oder war“ unterliegen (§ 88 Abs. 1 S. 1 TKG). Die den Datenschutz im Kern betreffenden Normen finden sich in Abschnitt 2. Normadressat sind „Unternehmen und Personen, die geschäftsmäßig Telekommunikationsdienste erbringen oder an deren Erbringung mitwirken“ (§ 90 Abs. 1 S. 2 TKG). Inhaltlich regelt das TKG insbesondere die Form der datenschutzrechtlichen Einwilligung (§ 94), den Umgang mit anfallenden Bestands (§ 95)-, Verkehrs (§ 96)- und Standortdaten (§ 98) sowie die Anforderungen an „Nachrichtenübermittlungssysteme mit Zwischenspeicherung[8]“ (§ 107).

Als weitere Regelung des Telekommunikationsrechts, die den Datenschutz tangiert, ist die Verordnung über die technische und organisatorische Umsetzung von Maßnahmen zur Überwachung der Telekommunikation (Telekommunikations-

[6] Selbst wenn man der Auffassung folgt, die Abgrenzung sei trennscharf anhand der Schichten des ISO/OSI- oder des TCP/IP-Modells zu treffen, so erscheint diese Grenzziehung doch nicht sachgerecht: Die systemunabhängige Darstellung von Daten, für die Schicht 6 zuständig ist, wird ebenso beispielsweise bei der reinen Sprachübertragung benötigt.

[7] Dennoch findet selbst bei solchen Protokollen *auch* Telekommunikation statt.

[8] Es handelt sich hierbei um Systeme, bei denen eine Nachricht zum Zweck eines späteren Abrufs (wie z.B. E-Mail) oder einer späteren Zustellung (z.B. beim Short Message Service in Mobilfunknetzen) zwischengespeichert wird.

Überwachungsverordnung, TKÜV) zu nennen; sie befasst sich mit Maßnahmen zur Telekommunikationsüberwachung durch Strafverfolgungsbehörden und Geheimdienste.

- *Schicht 2: Interaktion* Auf Schicht 2 wird die Interaktion eines Nutzers mit einem Dienste- bzw. Inhaltsanbieter betrachtet [Schl04, S. 729]. Auf Schicht 2 werden somit Inhalte der auf Schicht 1 stattfindenden Kommunikation betrachtet – allerdings nur, soweit die Kommunikation zwischen Nutzer und Dienst betroffen ist. Die Kommunikation zwischen Nutzer und Inhalteanbieter wird in dieser Schicht nicht betrachtet [Schl04, S. 729]. Typische Beispiele solcher Dienste sind Websites: Der Nutzer fordert mittels HTTP von einem Server Daten an. Der Webserver entscheidet aufgrund dieser Anfrage, welche Daten er ausliefern soll, und zeichnet gegebenenfalls Daten über das Nutzungsverhalten auf. Ob unter Verwendung des Dienstes, also beispielsweise eines Formulars auf der Website, Verträge geschlossen oder schutzwürdige Daten mit dem Inhalteanbieter ausgetauscht werden, ist hingegen in der Betrachtung auf Schicht 2 nicht relevant.

  Das Telemediengesetz[9] beinhaltet bereichsspezifische Regelungen zum Datenschutz; namentlich werden zusätzliche Pflichten der Diensteanbieter normiert (§ 13 TMG) sowie der Umgang mit Bestands (§ 14 TMG)-, Nutzungs (§ 15 TMG)- und Abrechnungsdaten (§ 15 Abs. 4–8 TMG) geregelt. Das Auskunftsrecht des Nutzers (§ 13 Abs. 7 TMG) und die Form der Einwilligung (§§ 12 Abs. 1–3, 13 Abs. 2 TMG) werden vom BDSG abweichend geregelt, indem jeweils die elektronische Form für zulässig erklärt wird. Soweit im TMG keine bereichsspezifischen Regelungen enthalten sind, gilt auch für den Datenschutz in Schicht 2 das BDSG.

- *Schicht 3: Inhalte* Auf Schicht 3 wird von sämtlichen technischen Gegebenheiten abstrahiert und nur noch der Inhalt der Kommunikation betrachtet [Schl04, S. 728]. Dementsprechend gelten auch keine telekommunikationsspezifischen gesetzlichen Regelungen, sondern nur das BDSG (und gegebenenfalls weitere Regelungen in Abhängigkeit von der Art des Kommunikationsinhalts). Wird beispielsweise ein Kaufvertrag unter Verwendung einer Website geschlossen, so ist für die Datenverarbeitung bei der Vertragsabwicklung lediglich das BDSG einschlägig, nicht aber TMG oder TKG. Wenig stichhaltig erscheint die Kritik von Imhof [Imho00, S. 113 ff.], der die Existenz von Inhaltsdaten bei Telediensten[10] abstreitet und davon ausgeht, das TDDSG[11] beziehe sich auch auf die im Rahmen der Nutzung eines Teledienstes übermittelten Daten. Dieser

[9] Dieses Gesetz führt die bislang in Teledienstegesetz (TDG), Teledienstedatenschutzgesetz (TDDSG) und Mediendienste-Staatsvertrag (MdStV) enthaltenen Regelungen zusammen. Ein Großteil dieser Regelungen wurde wörtlich aus den bisherigen Gesetzen übernommen. Für diese unveränderten Regelungen wird im Folgenden deshalb auch Literatur herangezogen, die sich noch auf TDG, TDDSG und MdStV bezieht.

[10] Seit Inkrafttreten des TMG: Telemedien.

[11] Diese Argumentation lässt sich auf das TMG übertragen.

Meinung nach unterfielen auch Daten, die im Rahmen einer Vertragsbeziehung (beispielsweise im Versandhandel) mit dem Nutzer ausgetauscht werden, dem TMG.

Einzelne Regelungen werden, soweit sie für das Szenario selbstorganisierender Empfehlungssysteme relevant sind, an geeigneter Stelle noch ausführlicher diskutiert.

Da, wie bereits erwähnt, das Recht alleine nicht zur Sicherstellung eines ausreichenden Datenschutzniveaus ausreicht und sogar selbst den Einsatz datenschutzfreundlicher Technik fordert (§ 3a BDSG), werden im nächsten Abschnitt Ansätze zum Datenschutz durch Technik diskutiert.

3.3 Technische Ansätze zum Datenschutz in Netzen

Techniken zur Verbesserung des Datenschutzniveaus werden auch als *Privacy Enhancing Technologies (PET)* bezeichnet. Solche Techniken lassen sich – in Abwandlung der Klassifikation von Grimm in [RoBG03, S. 82] – wie folgt klassifizieren:

- Werkzeuge zum Schutz der Identität des Nutzers (Anonymität und Pseudonymität) – sie lassen sich nach dem Modell der Nutzeranforderungen aus Abschnitt 3.1.1.1 der Identitätsebene zuordnen.
- Werkzeuge, die übermittelte Daten schützen (Integrität, Authentizität, Vertraulichkeit). Dienen sie der Vertraulichkeit, so lassen sie sich der Empfängerebene des Modells aus Abschnitt 3.1.1.1 zuordnen, da sie sicherstellen, dass nur beabsichtigte Empfänger Daten erhalten. Werkzeuge zum Schutz von Integrität und Authentizität lassen sich der Kontrollebene zurechnen, denn sie ermöglichen die Kontrolle darüber, ob die übermittelten Daten verändert wurden.
- Werkzeuge, die Transparenz über übermittelte Daten herstellen (sie betreffen die Inhalts- und die Kontrollebene):
  - Filter-Tools, die automatisch Daten ausfiltern, deren Übermittlung vom Nutzer nicht gewünscht wird. Transparenz wird dabei insofern hergestellt, als dem Nutzer die Kontrolle erleichtert wird, welche Daten übermittelt werden.
  - Policy-Tools, die Hinweise eines Datenempfängers bezüglich des Umgangs mit personenbezogenen Daten mit den Präferenzen des Nutzers abgleichen.
  - Sonstige Werkzeuge zur datenschutzbezogenen Kommunikation, die beispielsweise die datenschutzrechtliche Einwilligung des § 4a BDSG oder des § 4 Abs. 2 TDDSG erleichtern.

Eine weitere mögliche Klassifikation auf Basis des TDDSG (neu: des TMG) unterscheidet nach Transparenz, System- und Selbstdatenschutz (Grimm in [S. 83] [RoBG03]). Dabei

sollen Werkzeuge zur Untersützung der *Transparenz* dem Nutzer sichtbar machen, ob und wie mit seinen personenbezogenen Daten umgegangen wird. *Systemdatenschutz* bedeutet, dass Funktionen zum Datenschutz bereits dem System immanent sind, mit dem Daten verarbeitet werden. Gola/Klug [GoKl03, S. 47] bezeichnen dies als „Gestaltung der Systemstrukturen“ zur Reduktion der Erhebung, Verarbeitung und Nutzung personenbezogener Daten auf das notwendige Mindestmaß. Unter dem Begriff *Selbstdatenschutz* werden alle Datenschutzmaßnahmen zusammengefasst, die der Nutzer treffen kann – seien sie durch ein System unterstützt oder nicht.

Im Rahmen dieser Arbeit wird naturgemäß der Schwerpunkt auf Gesichtspunkten des Systemdatenschutzes liegen, da ein System entworfen wird, das personenbezogene Daten verarbeitet und in dem Datenschutzaspekte von Anfang an berücksichtigt werden. Dieser Schutz umfasst zum einen die Identität des Nutzers und zum anderen den Schutz der übermittelten Daten.

Im Folgenden werden einige bestehende Ansätze aus dem Bereich des Systemdatenschutzes klassifiziert und kurz vorgestellt. Dabei werden lediglich Ansätze berücksichtigt, die über den einfachen Einsatz von Verschlüsselungs- und Signaturverfahren zum Schutz von Vertraulichkeit, Integrität und Authentizität hinausgehen – auf diese wurde bereits in Abschnitt 2.1.2 eingegangen. Die betrachteten Verfahren setzen auf der Identitätsebene an, denn erst die Identität der Nutzer führt zu personenbezogenen Daten. Zugrunde liegende Idee aller Ansätze ist die Zusammenfassung mehrerer Nutzer zu einer Gruppe in der Art, dass der Kommunikationspartner oder ein Dritter lediglich die Gruppenzugehörigkeit des geschützten Nutzers, nicht aber seine genaue Identität feststellen kann. Die Qualität dieses Schutzes hängt somit von der Größe dieser Gruppe ab, die jedoch bei allen Verfahren variiert werden kann. Die vorgestellten Verfahren schützen lediglich den Telekommunikationsvorgang; ist die Identität eines Nutzers auf der Interaktions- oder Inhaltsebene verfügbar, kann ihr Schutz dadurch nicht gewährleistet werden.

3.3.1 Schutz der Identität durch Stellvertreter

Im einfachsten Fall wird die Identität eines Nutzers dadurch geschützt, dass seine Kommunikation über einen Stellvertreter abgewickelt wird. Alle Nachrichten werden zunächst an diesen gesendet, der sie an den jeweiligen Empfänger weiterleitet. Kommunikationspartner kennen lediglich die Identität des Stellvertreters. Das Prinzip kann im Internet auf verschiedenen Schichten angewendet werden. Auf Anwendungsschicht kann beispielsweise ein anonymisierende HTTP-Proxyserver verwendet werden, wie er zum Beispiel bei `http://www.anonymizer.com` oder `http://www.idzap.com` angeboten wird. Auf Vermittlungsschicht kann das Verfahren Network Address Translation (NAT) [RFC3022] eingesetzt werden. Hierbei ist für den Kommunikationspartner lediglich die IP-Adresse des Stellvertreters (Zwischensystems) sichtbar. In beiden Fällen ist die Gruppe, der ein Nutzer zugeord-

net werden kann, gleich der Gruppe aller Nutzer des jeweiligen Stellvertreters. Wesentlicher Nachteil der Lösungen ist das notwendige Vertrauen in den Stellvertreter bzw. dessen Betreiber: Diesem ist die Identität der an der Kommunikation beteiligten Knoten bekannt.

3.3.2 Schutz der Identität durch gruppeninterne Nachrichtenweiterleitung

Abhilfe können Verfahren schaffen, die mehrere Knoten zur Nachrichtenweiterleitung einsetzen. Solche Verfahren existieren in zahlreichen Varianten. In [Chau81] wird ein Verfahren für den anonymen Versand von E-Mails vorgeschlagen. Dabei wird die zu übertragende Nachricht m zunächst mit dem öffentlichen Schlüssel des Empfängers verschlüsselt. Die resultierende Nachricht m' wird zunächst einschließlich der Empfängeradresse mit dem öffentlichen Schlüssel eines Zwischensystems (der sogenannten *Mix*) verschlüsselt und dann an dieses System gesendet. Das Zwischensystem entschlüsselt Nachricht m' und Empfängeradresse und leitet die Nachricht dann an den eigentlichen Empfänger weiter. Um zu verhindern, dass durch den zeitlichen Ablauf bei Beobachtung des Mixes Rückschlüsse darauf gezogen werden können, welche aus- zu einer eingehenden Nachricht gehört, wird vor der Weiterleitung abgewartet, bis sich genügend Nachrichten beim Mix angesammelt haben.

Der Absender kann den benutzten Mix auswählen, insbesondere aber auch mehrere Mixes kaskadieren. In diesem Fall entschlüsselt der erste Mix die Adresse des zweiten Mix, kann jedoch das eigentliche Ziel nicht entschlüsseln. Der Absender kann einen beliebig langen Pfad durch verschiedene Mixes vorgeben, wobei jeder Mix nur seinen eigenen Vorgänger und Nachfolger sowie der Empfänger den letzten Mix im Pfad kennt. Solange nur einer der beteiligten Mixes die Beziehung zwischen ein- und ausgehender Nachricht geheimhält, ist die Identität des Absenders geschützt. Die Gruppe, der ein Nutzer zugeordnet werden kann, ist – sofern die Adressen der Nutzer öffentlich bekannt sind – bei diesem Verfahren identisch mit der Gesamtheit der Nutzer des Verfahrens.

Eine neuere Entwicklung, die jedoch auf dem gleichen Prinzip beruht, ist das *Onion Routing* [ReSG98]. Die Weiterleitung über mehrere Zwischensysteme erfolgt wie beim Mix-Ansatz, jedoch soll Onion Routing auch zum Einsatz kommen, wenn nur kurze Verzögerungen tolerabel sind. Aus diesem Grund wird auf die Zwischenspeicherung von Nachrichten verzichtet; stattdessen wird zufälliger, inhaltlich bedeutungsloser Netzwerkverkehr erzeugt, um die Verknüpfbarkeit ein- und ausgehender Nachrichten zu verhindern.

Der Begriff „Onion Routing“ rührt von der Analogie zu Zwiebelschalen. Jedes Zwischensystem entfernt eine Verschlüsselungsschicht und leitet die so „geschälte Zwiebel“ an das nächste System weiter. Auch hier kann, wie bei Mixes, ein Nutzer einer Gruppe zugeordnet werden kann, die – wiederum bei öffentlich bekannten Adressen der Nutzer – identisch mit der Gesamtheit der Nutzer des Verfahrens ist.

Ein weiterer vielbeachteter Ansatz zum Schutz der Identität eines Nutzers ist das *Crowds*-System [ReRu98]. Hier kann der Absender nicht den Pfad einer Nachricht festlegen. Statt-

dessen gilt innerhalb einer Gruppe (Crowd), dass jede Nachricht – sei es eine selbst generierte oder eine weiterzuleitende – mit einer gewissen Wahrscheinlichkeit p an das eigentliche Ziel gesendet, mit einer Wahrscheinlichkeit von $1 - p$ jedoch innerhalb der Gruppe weitergeleitet wird. Dieser Ansatz führt zwar dazu, dass der eigentliche Absender durch den Empfänger nicht identifiziert werden kann, doch ist die Erzeugung der Gruppen problematisch: [ReRu98, S. 7] schlägt vor, neue Gruppenmitglieder durch einen zentralen Server, den sogenannten Blender, autorisieren zu lassen. Die Notwendigkeit einer zentralen Instanz ist jedoch eine Schwachstelle, da alle Nutzer ihr vertrauen müssen.

Das Prinzip gruppeninterner Nachrichtenweiterleitung wird auch in anonymitätsunterstützenden Peer-to-Peer-Systemen eingesetzt; Beispiele sind Freenet [CSWH01], Tarzan [FSCM02] und [Serj02]. In der Regel wird dabei eine einzige Gruppe gebildet, in der alle am Peer-to-Peer-System teilnehmenden Knoten Mitglied sind. Die Verwaltung des Beitritts und des Verlassens der Gruppe wird durch das Peer-to-Peer-System übernommen.

3.4 Datenschutz in Peer-to-Peer-Systemen aus technischer Sicht

Konzepte wie das von Onion-Routing und Mixes können grundsätzlich auch in Peer-to-Peer-Systemen eingesetzt werden. Doch ändern sich durch den Einsatz von Peer-to-Peer-Technologie die Anforderungen an den Datenschutz: In reinen Web-Szenarien wird meist davon ausgegangen, dass derjenige, der Dienste in Anspruch nimmt oder Informationen abruft, schützenswert ist, der Anbieter jedoch nicht. Dies deckt sich auch mit der Rechtslage. In der Regel handelt es sich bei Websites um Telemedien[12]. Deren Anbieter sind ohnehin in vielen Fällen – bei „geschäftsmäßigen, in der Regel gegen Entgelt angebotenen Telemedien“ – zur Angabe von Namen, Anschrift und sonstigen Kontaktinformationen verpflichtet (§ 5 Abs. 1 TMG).

Wird ein Dienst jedoch in einem Peer-to-Peer-System erbracht, so ergeben sich aus technischer Sicht neue Möglichkeiten. So kann auch derjenige Nutzer, der Inhalte bereitstellt – und nicht nur derjenige, der sie abruft – seine Identität einfach verbergen. Wie dies im Einzelnen geschieht, wird in den folgenden Absätzen dargestellt. Inwieweit ein Verbergen der Anbieteridentität überhaupt wünschenswert und rechtlich zulässig ist, wird in Abschnitt 3.6.7.10 sowie in Kapitel 7 diskutiert.

3.4.1 Einfacher Fall: Inhaltsbasierte Adressierung

Auch ohne explizite Berücksichtigung von Datenschutzaspekten beim Entwurf von Peer-to-Peer-Systemen bleibt der Urheber eingespeister Inhalte dem abrufenden Knoten in vielen Fällen unbekannt. Dies wird beispielsweise bei der Adressierung von Dokumenten auf Basis ihres Inhalts, wie sie in Distributed Hashtables (DHTs) (beispielsweise CAN [RFHK+01]

[12] Zur Definition von Telemedien siehe noch Abschnitt 3.5.

und Chord [SMKK+01]) verwendet wird, erreicht. Dokumente werden in diesen Fällen nicht beim einspeisenden Knoten gespeichert, sondern bei dem Knoten, der für den aus dem jeweiligen Dokument ermittelten Hashwert zuständig ist. Wird nun ein solches Dokument abgerufen, so erfährt der abrufende Knoten lediglich die Adresse des speichernden Knotens, nicht aber die des Knotens, der den Inhalt eingespeist hat.

Dieses Prinzip alleine schützt zwar die Identität eines einspeisenden Knotens im Verhältnis zu demjenigen, der diese Inhalte abruft, doch bleiben zwei Probleme offen:

- Die Identität des Abrufenden wird nicht geschützt; dem oder den Knoten, die Inhalte speichern, ist bekannt, wer sich für diese interessiert.
- Die Identität des einspeisenden Knotens ist mindestens einem anderen Knoten bekannt.

Beide Probleme sind nicht für Peer-to-Peer-Systeme spezifisch; bei zentraler Diensterbringung können sie beispielsweise mit Hilfe von Onion Routing gelöst werden. Dies ist bei Peer-to-Peer-Systemen ebenso möglich, doch existieren daneben auch weitere Lösungsansätze. Exemplarisch soll an dieser Stelle das Freenet-System als bekanntester Ansatz kurz erläutert werden.

3.4.2 Freenet

Freenet [CSWH01] wurde mit dem Ziel entworfen, ein effizientes, vollständig verteiltes und zensurresistentes Peer-to-Peer-Filesharing-System zu schaffen, bei dem die Identität der Inhalteersteller und -abrufer geschützt ist und die speichernden Knoten die Kenntnis des Inhalts gespeicherter Dokumente abstreiten können.

Grundidee ist auch hier – wie bei Mixes, Onion Routing und Crowds – die Kombination der Weiterleitung von Nachrichten über mehrere Knoten mit der Verschlüsselung dieser Nachrichten. Dieses Prinzip wird sowohl beim Einfügen als auch beim Abrufen eines Dokuments verwendet. Dokumente werde in der Regel auf mehreren Knoten gespeichert. Sie werden vor dem Einfügen in das Netz zudem verschlüsselt – der speichernde Knoten kennt somit nicht notwendigerweise den Inhalt der bei ihm gespeicherten Daten. Zum einen führt dies dazu, dass das Nachverfolgen der Interessenten bestimmter Inhalte weiter erschwert wird. Zum anderen wird Zensur – das Ausfiltern unerwünschter Inhalte – auf diese Weise weitgehend verhindert. Gleichzeitig erschwert ein System, das konsequent auf die Anonymität seiner Teilnehmer ausgelegt ist, aber auch die Verfolgbarkeit von Rechtsverletzungen. Diese Problematik ist nicht auf Peer-to-Peer-Systeme beschränkt; durch den hohen Ressourcenaufwand, der bei zentraler Diensterbringung durch den Dienstgeber erbracht werden muss, ist der unauffällige Betrieb eines solchen rechtsverletzenden Systems in der Praxis jedoch wesentlich schwieriger und die praktische Verfolgbarkeit somit wesentlich eher gegeben. Die Problematik der Verfolgung von Rechtsverletzungen wird in Kapitel 7 diskutiert.

3.5 Rechtliche Einordnung selbstorganisierender Empfehlungssysteme

Die Einordnung selbstorganisierender Empfehlungssysteme in das Schichtenmodell des Telekommunikationsdatenschutzrechts (vgl. Abschnitt 3.2.2.1, S. 35) ist Grundvoraussetzung für die Beurteilung der sich ergebenden datenschutzrechtlichen Pflichten. Es stellt sich jedoch heraus, dass bereits die grundlegende Frage, ob und in welchem Umfang ein solches System dem Regelungsregime des Telemediengesetzes (TMG) unterworfen ist, erhebliche Probleme aufwirft. Deshalb wird zunächst diese Frage ausführlich diskutiert.

An dieser Stelle wird zunächst die Interaktionsebene betrachtet. Dazu werden als erstes klassische, auf dem Client-Server-Prinzip beruhende Empfehlungssysteme untersucht. Im Anschluss wird geprüft, inwieweit die gefundenen Ergebnisse im Fall eines Peer-to-Peer-Systems noch haltbar sind.

3.5.1 Erster Schritt: Zentralisierte Empfehlungssysteme

Es stellt sich die Frage, ob ein Empfehlungssystem ein Telemedium im Sinne des TMG darstellt. Da jeder bisherige Teledienst auch Telemedium ist[13], andererseits die Einordnung nach altem Recht noch für Altfälle relevant sein kann, bietet es sich an, die Definition von Telediensten nach TDG als Ausgangspunkt zu nehmen. § 2 I TDG definiert Teledienste als

> elektronische Informations- und Kommunikationsdienste, die für eine individuelle Nutzung von kombinierbaren Daten wie Zeichen, Bilder oder Töne bestimmt sind und denen eine Übermittlung mittels Telekommunikation zugrunde liegt.

Diese Definition ist von mehreren Seiten als unzureichend und zu wenig trennscharf kritisiert worden (siehe nur Gola/Müthlein in [GoMü00, S. 63, § 2 TDG Nr. 3.3]; Weinknecht in [Wein97], der den Charakter als Legaldefinition in Frage stellt). Dennoch soll der Wortlaut der Definition der Ausgangspunkt einer Prüfung der Anwendbarkeit des TDG auf selbstorganisierende Empfehlungssysteme sein, da er jedenfalls einen Rahmen für das Verständnis des Teledienstbegriffs vorgibt.

Dazu ist in einem ersten Schritt eine abstrakt-technische Zuordnung vorzunehmen (Tettenborn in [EFMT01, Rn. 41 zu § 2 TDG]):

- Handelt es sich um einen elektronischen Informations- und/oder Kommunikationsdienst? Die Definition des § 2 I TDG lässt aufgrund des verwendeten Plurals offen,

[13] Dies ergibt sich einerseits aus dem Wortlaut der Definitionen: Beide beinhalten den Begriff der „elektronischen Informations- und Kommunikationsdienste“, der jedoch im TMG nur gegen Telekommunikationsdienste, telekommunikationsgestützte Dienste und Rundfunk abgegrenzt, hingegen im TDG weiter eingegrenzt wird. Andererseits spricht dafür auch die Gesetzesbegründung des TMG [Bund06a, S. 11]

ob nur Dienste gemeint sind, die beide Aspekte (also Information und Kommunikation) beinhalten, oder ob einer dieser Aspekte ausreicht. Bereits die Bezeichnung als *Tele*dienst legt jedoch nahe, dass die Möglichkeit des Zugriffs aus der Ferne, also die Verwendung einer Telekommunikationsinfrastruktur vorausgesetzt wird. Wenn der Gesetzgeber nun die Begriffe Information und Kommunikation gleichwertig verwendet, so geht er wohl tatsächlich von einem Informations- *und* Kommunikationsdienst aus. So schlägt auch Tettenborn [EFMT01, Rn. 46 zu § 2 TDG] als erstes Prüfkriterium die Frage vor, ob es sich um ein „elektronisch gespeichertes, mittels Telekommunikation individuell und unmittelbar (ohne Medienbruch) abrufbares inhaltliches Angebot" handelt. Der Begriff eines „gespeicherten" Angebots ist jedoch eine zu weitgehende Einschränkung, da auch ein Angebot, das in dieser Form erst auf Nutzeranfrage automatisch generiert, jedoch nur zum Zweck der Übermittlung kurzfristig zwischengespeichert wird, als Informations- und Kommunikationsdienst zu qualifizieren ist: Zumindest sind dem Gesetz keine gegenteiligen Anhaltspunkte zu entnehmen. Eine Unterscheidung anhand dieses, für den Anwender kaum überprüfbaren Kriteriums wäre auch nicht sachgerecht.

In der Praxis dürfte bei einem Empfehlungssystem immer von einem Informations- und Kommunikationsdienst auszugehen sein: Das Angebot kommt zwar erst durch Beiträge der Nutzer zustande, die explizit oder implizit Bewertungen abgeben oder doch zumindest in einer Weise handeln, die dem System die Ableitung von Empfehlungen ermöglicht. Doch dies ist für die rechtliche Bewertung des Abrufs nicht relevant. Die Empfehlungen oder doch zumindest die Grundlagen, aufgrund derer sie berechnet werden, sind auf einem Rechnersystem des Systembetreibers gespeichert. Ein Nutzer, der dem System eine Empfehlung entnehmen will, ruft diese mit Mitteln der Telekommunikation (in der Regel unter Verwendung des Internets) individuell und ohne Medienbruch ab. Zu bemerken gilt hier noch, dass Telekommunikation eingesetzt wird, das Empfehlungssystem aber selbst keinen Telekommunikationsdienst darstellt, denn der Dienst besteht nicht „ganz oder überwiegend in der Übertragung von Signalen über Telekommunikationsnetze" (§ 3 Nr. 24 TKG).

- Aus den Ausführungen des letzten Absatzes ergibt sich auch bereits, dass dem Dienst eine „Übermittlung mittels Telekommunikation zugrunde liegt". Auch dieses Kriterium des § 2 I TDG kann also als erfüllt angesehen werden.

- Werden „kombinierbare Daten wie Zeichen, Bilder oder Töne" genutzt? Es sind verschiedene Präsentationsformen einer Empfehlung denkbar, doch ist davon auszugehen, dass jede dieser Formen zumindest auch auf Zeichen, Bilder oder Töne zurückgreift. Nicht zu folgen ist Tettenborn (in [EFMT01, Rn. 41, 46]), der eine „multimediale Nutzung" verlangt und von „monomedialen Diensten bzw. Angeboten" abgrenzen

will. Auch reine Textdarstellungen können Teledienste sein. Weder der Wortlaut des Gesetzes noch dessen amtliche Begründung [Bund97] sehen die Integration mehrerer Medien als zwingende Voraussetzung für das Vorliegen eines Teledienstes.

- Sind diese Daten auch für die individuelle Nutzung bestimmt? Zumindest aus rein technischer Sicht lässt sich dies bejahen: Die Daten werden an jeden Nutzer individuell übermittelt; sie können sogar beim Abruf durch einen Nutzer erst für diesen erzeugt werden, was jedoch für das Vorliegen eines Teledienstes keine Voraussetzung ist.

Aus „technisch-abstrakter“ [EFMT01, Rn. 46] Sicht kann also für Empfehlungssysteme die Eigenschaft als Teledienste bejaht werden. Dem Prüfungsschema aus [EFMT01, Rn. 46] folgend, schließt sich jedoch eine funktionale Prüfung an. In deren erstem Schritt gilt es festzustellen, ob sich Empfehlungssysteme in die Liste der Teledienste des § 2 II TDG einordnen lassen. In Frage kommen hier lediglich zwei Elemente dieser Liste; als Teledienste eingeordnet werden demnach

- „Angebote im Bereich der Individualkommunikation (zum Beispiel Telebanking, Datenaustausch)“ (§ 2 II Nr. 1 TDG). Ein Empfehlungsdienst beschränkt sich jedoch in aller Regel nicht darauf, den Austausch von Empfehlungen zwischen individuellen Nutzern zu ermöglichen. Vielmehr findet eine Aggregation und eine Speicherung von Bewertungen statt. Empfehlungsdienste unterfallen also nicht dem § 2 II Nr. 1 TDG.

- „Angebote zur Information oder Kommunikation, soweit nicht die redaktionelle Gestaltung zur Meinungsbildung für die Öffentlichkeit im Vordergrund steht (Datendienste, zum Beispiel Verkehrs-, Wetter-, Umwelt- und Börsendaten, Verbreitung von Informationen über Waren und Dienstleistungsangebote)“ (§ 2 II Nr. 2 TDG). In der Tat ist ein Empfehlungsdienst ein „Angebot zur Information“; dem Nutzer werden Informationen darüber angeboten, welche Objekte für ihn interessant sein könnten bzw. wie andere Nutzer ihn interessierende Objekte bewerten. Auch wird die „Verbreitung von Informationen über Waren und Dienstleistungsangebote“ explizit aufgeführt. Zwar lässt die Gesetzesbegründung [Bund97, S. 19] darauf schließen, dass der Gesetzgeber damit ursprünglich Angebote zum Zweck der Werbung meinte. Doch spricht nichts dagegen, auch unabhängige Produktinformationen entsprechend einzuordnen.

  Auch dem Charakter als Angebot steht nichts entgegen – insbesondere liegt ein Anbieter-Nutzer-Verhältnis vor, bei dem der Betreiber des Empfehlungsdienstes der Anbieter ist. Vorläufig sind Empfehlungsdienste also als Teledienste im Sinne des TDG einzuordnen.

  Problematisch könnte allein die Einschränkung sein, nach der die „redaktionelle Gestaltung zur Meinungsbildung für die Öffentlichkeit“ nicht im Vordergrund stehen

darf. Sie dient der Abgrenzung zu Mediendiensten, der im Folgenden ein eigener Absatz gewidmet wird.

3.5.1.1 Abgrenzung zu Mediendiensten

Die Existenz eines Teledienstegesetzes und eines Mediendienstestaatsvertrags (MdStV), deren Anwendungsbereich zumindest in Teilgebieten schwer abgrenzbar[14] war, liegt in den Gesetzgebungskompetenzen von Bund und Ländern begründet, die zu einer Zeit geregelt wurden, als die zunehmende Medienkonvergenz noch nicht absehbar war. Im Rahmen eines 1996 geschlossenen Kompromisses einigten sich Bund und Länder auf die Schaffung eines in der Sache einheitlichen Rechtsrahmens in TDG und MdStV [Sieb99, S. 3]. In wesentlichen Punkten, insbesondere den Regelungen über Verantwortlichkeiten für bereitgehaltene oder übermittelte Inhalte, stimmten beide Regelungswerke inhaltlich überein. Doch gibt es auch Abweichungen, die teils durch die Nähe von Mediendiensten zur Presse bedingt sind (beispielsweise der Anspruch auf Gegendarstellung aus § 14 MdStV, der im TDG kein Pendant fand). Im Detail unterschiedlich waren auch die Regelungen zur Anbieterkennzeichnung [Brun04, S. 8]. Eine Abgrenzung war also zumindest nicht in allen Fällen verzichtbar.

Abgrenzungskriterium war bereits nach dem Wortlaut der gesetzlichen Regelungen (§ 2 I Satz 1 MdStV, § 2 I und § 2 IV Nr. 3 TDG) die Frage, ob „die redaktionelle Gestaltung zur Meinungsbildung für die Allgemeinheit“ (§ 2 IV Nr. 3 TDG) im Vordergrund steht. Leitbild sind hier Presse und sonstige Druckerzeugnisse (Tettenborn in [EFMT01, Rn. 46 zu § 2 TDG]).

Wie ist nun ein Empfehlungsdienst einzuordnen? Wesentliche Aufgaben eines solchen Dienstes sind

- einem Nutzer das Auffinden potentiell interessanter Objekte zu ermöglichen;
- einem Nutzer Bewertungen von Objekten, die ihm (durch das System selbst oder auf anderem Weg) bekannt sind, zur Verfügung zu stellen.

Beide Funktionen sind grundsätzlich im Rahmen eines Angebots denkbar, bei dem „die redaktionelle Gestaltung zur Meinungsbildung für die Allgemeinheit“ im Vordergrund steht. Ein solches Angebot könnte die elektronische Variante eines Warentest-Magazins darstellen, in dem Produkte vorgestellt und bewertet werden. Jedoch wäre ein solches Angebot nur im weiteren Sinne als Empfehlungssystem zu bezeichnen, denn aufgrund der fehlenden Personalisierung kann es eben keine maßgeschneiderten Empfehlungen geben.

Werden einem Nutzer aufgrund seines Profils jedoch personalisierte Empfehlungen angeboten, wird man im allgemeinen nicht mehr davon sprechen können, dass die redaktionelle

[14] Siehe nur [Wald98, S. 124], [Goun97, S. 2994].

Gestaltung im Vordergrund stehe. Dies ist lediglich dann noch der Fall, wenn auf Grundlage eines Profils ein redaktionell aufbereitetes Angebot angezeigt wird. Auch dann gilt eine Empfehlung jedoch nicht mehr für die Allgemeinheit, kann also auch schwerlich zur „Meinungsbildung für die Allgemeinheit“ beitragen.

Zwar sind Ausnahmen denkbar, beispielsweise wenn der Empfehlungsdienst nur als Teil einer anderen Dienstleistung erbracht wird, die selbst einen Mediendienst darstellt; in diesem Fall ist, so man noch von einem einheitlichen Dienst sprechen kann, eine „Gesamtschau“ vorzunehmen (so [Wald98, S. 125]; anderer Auffassung [Tett99, S. 518][15]). Doch in der Regel werden (zentralisierte) Empfehlungssysteme aus den genannten Gründen als Teledienste zu qualifizieren sein. In der Literatur wird die Meinung vertreten, dass ohnehin im Zweifel von einem Teledienst auszugehen sei, da Bundesrecht (also auch das TDG) nach Artikel 31 GG Landesrecht bricht (so Waldenberger in [Wald98, S. 126], Tettenborn in [EFMT01, vor § 1 TDG Rn. 20]; anderer Auffassung Spindler in [SpSG04, Rn. 21 zu § 2 TDG]). Für den konkreten Fall der Empfehlungssysteme kann dahinstehen, ob dieser Ansicht gefolgt werden kann. Im Ergebnis bleibt festzuhalten, dass jedenfalls ein zentralisiertes Empfehlungssystem in aller Regel als Teledienst einzuordnen war. Nach neuem Recht fällt diese Abgrenzung weg, denn Mediendienste-Staatsvertrag und Teledienstegesetz wurden im TMG zusammengeführt. Zentralisierte Empfehlungssysteme sind somit – da die Definition der Telemedien in § 1 Abs. 1 TMG sowohl Teledienste als auch Mediendienste nach altem Recht umfasst – in jedem Fall als Telemedien einzuordnen.

3.5.2 Auswirkungen der Eigenschaften von Peer-to-Peer-Systemen

Dieses Ergebnis nun auf ein verteiltes, selbstorganisierendes Empfehlungssystem zu übertragen, erscheint auf den ersten Blick trivial. Die erbrachte Funktion ändert sich nicht oder nur marginal. Im Extremfall kann durch die Nutzerschnittstelle sogar völlig von der technischen Funktionsweise abstrahiert werden.

Doch auf der anderen Seite ändert sich die technische Funktionsweise, also die Art, in der der Dienst erbracht wird, vollkommen: Es gibt nicht mehr den einen Anbieter, an den der Nutzer sich wenden kann. Doch rechtfertigt dies eine gänzlich andere Behandlung als im Fall des zentralisierten Systems? Um dies zu beantworten, müssen lediglich zwei Fragen gestellt werden, die eng miteinander verknüpft sind: Liegt nach wie vor ein Anbieter-

[15] Der Widerspruch beider Autoren liegt jedoch vornehmlich in der Frage, wann noch ein einheitlicher Dienst vorliegt; Waldenberger [Wald98, S. 125] geht davon aus, dass das „vollständige, unter einer bestimmten Homepage (Eingangsseite) einschließlich der untergeordneten Seiten abrufbare Angebot eines Unternehmens“ als ein einheitlicher Dienst anzusehen ist und geht damit deutlich zu weit. Vielmehr ist von einem einheitlichen Dienst nur dann auszugehen, wenn sich mehrere vorhandene Angebote „als Einheit darstellen, da sich dem Nutzer nicht die unterschiedlichen Funktionen der Individual- und der Massenkommunikation aufdrängen“ (so Spindler in [SpSG04, Rn. 39 zu § 2TDG]). Zum Begriff des einheitlichen Dienstes vgl. auch Abschnitt 3.6.7.5

Nutzer-Verhältnis vor? Und worin liegt eigentlich bei einem selbstorganisierenden, dezentralen System der erbrachte Dienst?

3.5.2.1 Anbieter-Nutzer-Verhältnis

Tettenborn [Tett99, S. 518] weist darauf hin, ein Teledienst könne – wie in § 2 Abs. 2 TDG deutlich werde – nur vorliegen, wenn ein Angebot bestehe, also ein Anbieter-Nutzer-Verhältnis vorliege (so auch Gola/Müthlein in [GoMü00, S. 88, § 2 TDG Nr. 6.4.1]). Nun enthält § 2 Abs. 2 TDG lediglich Regelbeispiele („insbesondere") für Teledienste, aber keine abschließende Aufzählung. Die Tatsache, dass alle fünf Regelbeispiele den Begriff „Angebot" beinhalten, ist ein Indiz, dass der Gesetzgeber tatsächlich nur diesen Fall im Blick hatte. Jedenfalls kann aber der von Tettenborn in [EFMT01, Rn. 40 zu § 2 TDG] vertretenen Auffassung gefolgt werden, nach der der in § 2 Abs. 1 TDG verwendete Begriff „Dienst" als Synonym für „Angebot"[16] verwendet wird. Im Ergebnis kann ein Teledienst also nur bestehen, wenn ein Angebot, also auch ein Anbieter vorhanden ist. Ein „Nutzer-Nutzer-Verhältnis" reicht nicht aus [EFMT01, Rn. 40 zu § 2 TDG]. Als Beispiele werden firmeninterne Kommunikation [Tett99, S. 518] und der Aufbau von Punkt-zu-Punkt-Verbindungen mittels ISDN (Gola/Müthlein in [GoMü00, S. 75, § 2 TDG Nr. 4.2.3]) herangezogen. In beiden Fällen wird kein Dienst erbracht (bzw. es liegt kein Angebot vor). Diese in der Literatur herangezogenen Beispiele sind missverständlich; die Herstellung einer Verbindung mittels ISDN betrifft nicht die Interaktionsebene, und prinzipiell kann auch auf diese Weise die Nutzung eines Teledienstes ermöglicht werden. Der Interpretation, dass bei gleichberechtigter Kommunikation unter Nutzern kein Teledienst vorliegt, kann dennoch zugestimmt werden. Dies gilt auch für Telemedien: Dass ein Anbieter-Nutzer-Verhältnis für die Geltung der datenschutzrechtlichen Normen des TMG erforderlich ist, hat der Gesetzgeber für zwei Spezialfälle (Dienst- und Arbeitsverhältnis, § 11 Abs. 1 Nr. 1 TMG sowie die Steuerung von Arbeits- oder Geschäftsprozessen in § 11 Abs. 1 Nr. 2 TMG) explizit klargestellt. Die weitere Verwendung des Begriffs „Informations- und Kommunikationsdienst", der Hinweis des Gesetzgebers, TDG und MdStV zusammenführen zu wollen [Bund06a, S. 13] und die Beibehaltung des Begriffs der „Abhol- und Verteildienste" (ebenda) deuten darauf hin, dass in dieser Hinsicht auch jenseits der genannten Spezialfälle keine Änderung beabsichtigt war.

Doch wie genau ist zwischen Kommunikation unter Nutzern und der Erbringung eines Teledienstes bzw. Telemediums abzugrenzen? Eine Möglichkeit wäre, wie durch den Begriff „Nutzer-Nutzer-Verhältnis" nahegelegt, einen Dienst stets dann abzulehnen, wenn zwei gleichberechtigte Teilnehmer beteiligt sind. In dieser Allgemeinheit wäre das jedoch zu kurz gegriffen. Vielmehr gilt zu differenzieren: Bearbeiten diese Teilnehmer *gemeinsam* eine Aufgabe zum gegenseitigen Nutzen (oder zum Nutzen eines Dritten), so kann keine Partei

[16] Der Angebotsbegriff ist hier jedoch nicht im Sinne des Vertragsrechts als Willenserklärung zu verstehen, sondern folgt dem allgemeinen Sprachgebrauch.

(nur) als Dienstnehmer oder Dienstgeber identifiziert werden. In diesem Fall liegt also auch im Verhältnis zwischen den Teilnehmern kein Teledienst vor. Wenn jedoch jede Seite die andere mit der Erfüllung von Aufgaben beauftragen kann, so handelt es sich um einen Teledienst, bei dem lediglich die ständige Möglichkeit zum Rollenwechsel besteht. Es besteht kein Anlass, einem Nutzer, der selbst auch Dienste erbringt, deshalb den Schutz des Teledienstegesetzes und des Teledienstedatenschutzgesetzes bzw. des Telemediengesetzes zu nehmen. Dennoch bleibt die Abgrenzung schwierig, da der Übergang von der echten Zusammenarbeit zur Erbringung eines Dienstes mit Möglichkeit zum Rollenwechsel fließend ist. Im konkreten Fall selbstorganisierender Empfehlungssysteme ist jedoch davon auszugehen, dass die Aufgabe „Erstellen einer *konkreten* Objektempfehlung“ eben nicht von beiden Seiten nachgefragt wird. Dies gilt zunächst unabhängig davon, ob als Gegenüber des Nutzers ein einzelner Knoten oder das System als Ganzes gesehen wird. Welche Probleme sich in beiden Fällen ergeben können, wird im nächsten Abschnitt diskutiert.

3.5.2.2 Erbrachter Dienst

Grundsätzlich kann also auch in einem Peer-to-Peer-System ein Dienst erbracht werden. Es stellt sich jedoch die Frage, worin genau der Dienst gesehen werden kann: Viele – potentiell alle – Netzteilnehmer wirken zusammen. Sie erledigen Teilaufgaben, wie das Weiterleiten von Anfragen oder das Versenden von Bewertungen, um die gewünschten Gesamtaufgabe (einem Nutzer ein Objekt zu empfehlen) zu erbringen. Liegt nun der Teledienst bereits in der Bearbeitung der Teilaufgaben, bietet also jeder der Teilnehmer einen Dienst an? Oder liegt ein Teledienst bzw. Telemedium erst bei Betrachtung des Gesamtsystems vor, ist also das Erfüllen der „Gesamtaufgabe“ erst der Dienst? Beide Möglichkeiten sollen im Folgenden erörtert werden.

Teilaufgaben als Teledienste bzw. Telemedien. Die von den einzelnen Knoten auszuführenden Teilaufgaben lassen sich im wesentlichen in zwei Bereiche unterteilen: zum einen die *Beantwortung* von Anfragen, beispielsweise durch Versenden einer Objektbewertung (sei es aufgrund lokal gespeicherter Bewertungen oder aufgrund von Informationen, die andere Knoten beigesteuert haben), zum anderen das reine *Weiterleiten* solcher Anfragen.

- Was die Beantwortung von Anfragen angeht, so ändert sich an der Argumentation zur Einordnung als Teledienst gegenüber dem bereits diskutierten zentralisierten Empfehlungssystem zunächst nichts; es bleibt bei einem Angebot der „Verbreitung von Informationen über Waren und Dienstleistungsangebote)“ (§ 2 II Nr. 2 TDG). Doch stellt sich die Frage, ob dieses Angebot umfassend genug ist, um als Teledienst betrachtet zu werden: Es ist gut denkbar, dass auf einem Knoten nur eine einzige Objektbewertung gespeichert ist. Ein beliebiger Nutzer wird also einen bestimmten Knoten nur mit einer geringen Wahrscheinlichkeit überhaupt in Anspruch nehmen. Wenn er dies

tut, wird er in der Regel Informationen von mehreren Knoten verwenden; der Beitrag eines einzelnen Knotens zur letztendlich errechneten Empfehlung ist also zumindest potentiell eher gering. Es kann somit nur von einem wenig umfangreichen Angebot ausgegangen werden, dessen Qualifikation als Teledienst es noch zu diskutieren gilt. Dies hängt eng damit zusammen, dass ein Teilnehmer primär nicht die Dienstleistung im Blick hat, die ein einzelner Knoten anbietet, sondern nur am Gesamtergebnis interessiert ist.

All dies spricht auf den ersten Blick dagegen, die Erbringung eines Teledienstes bzw. Telemediums durch die einzelnen Knoten anzunehmen. Einer näheren Betrachtung halten diese Argumente jedoch nicht stand. Dass ein einzelner Knoten nur mit geringer Wahrscheinlichkeit überhaupt durch einen Nutzer in Anspruch genommen wird, ändert nichts daran, dass im Fall der Inanspruchnahme ein Teledienst vorliegen kann. Die Situation ist der Inanspruchnahme einer Website, wie beispielsweise einer privaten Homepage, die in aller Regel ein Telemedium bzw. einen Teledienst (oder bei redaktioneller Gestaltung einen Mediendienst) darstellt (vgl. nur Tettenborn in [EFMT01, Rn. 53 zu § 2 TDG], [EFMT98, S. 11]), nicht unähnlich. Auch der geringe Umfang der erbrachten Dienstleistung – der aus Sicht des Nutzers zu beurteilen ist – ist unschädlich: So wird in der Literatur bereits die Vergabe von IP-Adressen (beispielsweise mittels des Dynamic Host Configuration Protocol [Drom97]) als Teledienst eingeordnet (so z.B. im Evaluierungsbericht der Bundesregierung zum Informations- und Kommunikationsdienstegesetz, [Bund99, S. 7 f.]). Zwar ist diese Einordnung abzulehnen, aber nur, da die Vergabe von IP-Adressen nicht auf der Inhalteebene stattfindet, sondern der technischen Funktion des Netzes zuzuordnen und mithin nach TKG zu beurteilen ist (so auch [LGDA05, Abs. 41]; Spindler in [SpSG04, Rn. 26 zu § 2 TDG]). Der Umfang dieses Dienstes wird in der betrachteten Literatur jedoch durchweg nicht als problematisch angesehen. Wird DHCP lediglich zur Vergabe von IP-Adressen eingesetzt, so wird dieser Dienst dem Nutzer im Normalfall nicht einmal bewusst. Das Beispiel zeigt deutlich, dass es praktisch keine untere Schranke für Umfang und Komplexität eines Teledienstes gibt.

- Betreffend die reine Weiterleitung von Anfragen ergibt sich noch eine weitere Abgrenzungsfrage. Das Teledienstegesetz gilt nicht für Telekommunikationsdienstleistungen (§ 2 Abs. 4 Nr. 1 TDG). Dieser in der alten Fassung des Telekommunikationsgesetzes (TKG) definierte Begriff wurde durch den Begriff des „Telekommunikationsdienstes“ ersetzt, was – trotz abweichender Definition – für den konkreten Anwendungsfall jedoch nicht von Belang ist. Das Telemediengesetz nimmt (in § 1 Abs. 1) bereits bezug auf den neuen Begriff des Telekommunikationsdiensts. in Telekommunikationsdienste sind „in der Regel gegen Entgelt erbrachte Dienste, die ganz oder überwiegend in der Übertragung von Signalen über Telekommunikationsnetze bestehen [...]“ (§ 3 S. 1

Nr. 24 TKG). Aus zweierlei Gründen trifft diese Einschränkung des Anwendungsbereichs des TDG bzw. des TMG jedoch für selbstorganisierende Empfehlungssysteme nicht zu:

- Zum einen werden Anfragen in Peer-to-Peer-Systemen in der Regel nicht gegen Entgelt[17] weitergeleitet; eine entsprechende Entwicklung ist auch nicht absehbar. Selbst wenn ein einzelnes System vorsieht, eine Entgeltlichkeit der Weiterleitung von Anfragen einzuführen, so ist damit noch kein „Regelfall" begründet und somit noch nicht von einem Telekommunikationsdienst auszugehen. Die Entgeltlichkeit der Erbringung anderer Dienste, die nicht lediglich in der Weiterleitung von Anfragen bestehen, ist davon unabhängig.
- Zum anderen handelt es sich in aller Regel nicht um eine reine Weiterleitung an einen vorher bestimmten Adressaten, sondern der weiterleitende Knoten wählt den oder die Empfänger anhand inhaltlicher Kriterien der Nachricht selbst aus[18]. Folgt man der in der Literatur überwiegend vertretenen funktionalen Abgrenzung (siehe nur Neubauer in [MoDr05, Teil D, Rn. 6], Spindler in [SpSG04, Rn. 22, 26 zu § 2 TDG]), nach der inhaltsbezogene Dienste dem TDG (neu: dem TMG) zuzuordnen sind, die rein technische Dienstleistung jedoch dem TKG, so führt dies zur Einordnung als Teledienst (neu: als Telemedium) und nicht als Telekommunikationsdienst.[19] Auch wird zwar argumentiert, die Anwendbarkeit des TDG setze eine unmittelbare Kommunikationsbeziehung zum Nutzer des Dienstes voraus, was sich im Katalog von Telediensten des § 2 Abs. 2 TDG zeige[20] (Spindler in [SpSG04, Rn. 31 zu § 2 TDG]); diese Argumentation verkennt aber, dass der Gesetzgeber das TDG bewusst für neue Entwicklungen offenhalten wollte [Bund97] und kein Grund besteht, die Rechtsstellung des Anbieters einer Dienstleistung von der Unmittelbarkeit der Kommunikation abhängig zu machen. Vielmehr genügt die funktionale Abgrenzung, die bezüglich der Einordnung von Routern, wie sie durch Spindler in [SpSG04, Rn. 31 zu § 2 TDG] diskutiert wird, ohnehin zum gleichen Ergebnis führt.

Insgesamt ist also auch die Weiterleitung von Anfragen im Rahmen des Empfehlungssystems als Teledienst gemäß § 2 Abs. 1 TDG bzw. als Telemedium gemäß § 1 Abs. 1 TMG einzuordnen.

[17] Der Begriff des Entgelts ist im TKG nicht definiert. Es ist davon auszugehen, dass ein Entgelt nicht in Geld bestehen muss; wenn kein durchsetzbarer Anspruch auf eine Gegenleistung erworben wird, sondern lediglich Anreize zu einem bestimmten Verhalten geschaffen werden, kann indes nicht von einem Entgelt die Rede sein.

[18] Dies gilt, auch wenn der weiterleitende Knoten in dieser Auswahl nicht völlig frei ist, sondern anhand eines vorgegebenen Protokolls entscheidet.

[19] Anders stellt sich die Sachlage bei Routern dar, die den Inhalt von Datenpaketen gar nicht zur Kenntnis nehmen müssen, sondern lediglich auf Grundlage von IP-Adressen Pakete weiterleiten können (anderer Auffassung [Roßn00, S. 625]).

[20] Diese Argumentation ist mit Wegfall dieses Katalogs im TMG ohnehin überholt.

Vorläufig kann als Ergebnis also festgehalten werden, dass die Teilnehmer eines selbstorganisierenden Empfehlungssystems in aller Regel einen Teledienst erbringen (so für den Fall von Peer-to-Peer-Filesharing-Netzen ohne weitere Diskussion auch [Koch05, S. 752]). Eingeschränkt werden könnte dieses Ergebnis jedoch, falls sich herausstellen sollte, dass

- das Gesamtsystem als Teledienst einzuordnen ist
- und dies die Teledienstеigenschaft seiner Bestandteile verhindert. Dies wäre der Fall, falls ein Angebot nicht gleichzeitig Teledienst und Teil eines umfassenderen Teledienstes sein könnte.

Diese Fragestellungen werden im Folgenden diskutiert.

Gesamtsystem als Teledienst Die Frage, ob ein Peer-to-Peer-System als Ganzes als Teledienst bzw. Telemedium zu qualifizieren ist, hat in der Literatur bisher nur am Rande Erwähnung gefunden. Brinkel [Brin06, S. 263] bezeichnet die Subsumtion von Peer-to-Peer-Systemen unter die Kategorien des TDG als problematisch bis unmöglich; Wenzl [Wenz05, S.124] argumentiert (ebenfalls von Filesharing-Systemen ausgehend), dass sich bei „echten" Peer-to-Peer-Systemen kein Diensteanbieter identifizieren lasse und daher kein „Dienst" im Sinne des TDG vorliege. Wie bereits angedeutet, erfüllt das Gesamtsystem die gleichen Aufgaben wie ein zentralisiertes Empfehlungssystem. Aus funktionaler Sicht ist das System also als Teledienst zu qualifizieren. Auch ein Anbieter-Nutzer-Verhältnis könnte grundsätzlich vorliegen (s. oben S. 49). Probleme ergeben sich jedoch aus zweierlei Gründen: Praktisch ist es nicht möglich, einen Anspruch gegen ein dezentrales, selbstorganisierendes System durchzusetzen. Eng damit zusammen hängt der zweite Grund, nämlich die Definition des Diensteanbieters in § 3 Satz 1 Nr. 1 TDG bzw. § 2 Satz 1 Nr. 1 TMG. Diensteanbieter ist demnach

> jede natürliche oder juristische Person, die eigene oder fremde Teledienste[21] zur Nutzung bereithält oder den Zugang zur Nutzung vermittelt.

Diejenigen Personengesellschaften, die mit der Fähigkeit ausgestattet ist, Rechte zu erwerben und Verbindlichkeiten einzugehen, sind gemäß § 3 Satz 2 TDG (§ 2 Satz 2 TMG) juristischen Personen gleichgestellt.

Zur rechtlichen Beurteilung des Systems sind somit drei Fragen zu beantworten:

- Lässt sich eine natürliche Person, eine juristische Person oder eine teilrechtsfähige Personengesellschaft identifizieren, die als Anbieter des Gesamtsystems betrachtet werden kann?

[21] In § 2 Satz 1 Nr. 1 TMG: Telemedien

- Falls dies nicht der Fall ist, kann das Gesamtsystem dennoch ein Teledienst bzw. Telemedium sein, ohne dass ein Diensteanbieter im Sinne des § 3 S. 1 Nr. 1 TDG (§ 2 Satz 1 Nr. 1 TMG) vorliegt? Da die weiteren Normen des TDG bzw. des TMG an Diensteanbieter gerichtet sind, ist diese Frage von geringer praktischer Relevanz. Dennoch soll sie an dieser Stelle kurz diskutiert werden.

- Falls das Gesamtsystem ein Teledienst ist, können dann die einzelnen Knoten des Systems nach wie vor als Diensteanbieter eingeordnet werden?

Diese Fragen werden in den folgenden Absätzen diskutiert.

Person oder Personengesellschaft als Betreiber? In Frage kommen gemäß § 3 S. 1 Nr. 1, § 3 S. 2 TDG (§ 2 Satz 1 Nr. 1, § 2 Satz 2 TMG) nur natürliche Personen, juristische Personen und teilrechtsfähige Personengesellschaften. Lässt sich nun eine dieser drei Alternativen als Betreiber selbstorganisierender Systeme identifizieren? Bevor die Betreibereigenschaft desjenigen diskutiert wird, der das System entwirft und die zu seinem Betrieb notwendige Software bereitstellt, sollen zunächst die Systemteilnehmer selbst betrachtet werden.

- Eine natürliche Person lässt sich nicht als Systembetreiber identifizieren, da das System sich gerade durch die Mitwirkung einer Vielzahl natürlicher Personen auszeichnet.

- Das Vorliegen einer *juristischen Person* scheidet aus, da juristische Personen nach deutschem Recht ausschließlich durch Registereintragung, behördliche Anerkennung oder staatliche Verleihung entstehen können (vgl. Heinrichs in [Pala06, Rn. 6,7 vor § 21]).

- Denkbar wäre jedoch das Vorliegen einer teilrechtsfähigen Personengemeinschaft. In Frage kommen der nichtrechtsfähige Verein[22] und die Gesellschaft bürgerlichen Rechts.

  – Der *nichtrechtsfähige Verein* ist in § 54 BGB geregelt; demnach finden die Vorschriften über die Gesellschaft bürgerlichen Rechts Anwendung (§ 54 S. 1 BGB). Rechtsprechung und Literatur wenden jedoch die Vorschriften über den rechtsfähigen Verein analog an, was mit dem zumindest stillschweigenden Abbedingen der unpassenden Regelungen des Gesellschaftsrechts begründet wird [Brox05, Rn. 770]. Ein nichtrechtsfähiger Verein ist eine auf Dauer angelegte und auf die Erreichung eines gemeinsamen Zwecks gerichtete Verbindung einer größeren Personenzahl, die „nach ihrer Satzung körperschaftlich organisiert ist, einen Gesellschaftsnamen führt und auf einen wechselnden Mitgliederbestand angelegt ist“ (Heinrichs in [Pala06, Rn. 1 zu § 54]). Der nichteingetragene[23] Verein ist teilrechtsfähig (Heinrichs in [Pala06, Rn. 7 zu § 54]). Doch lassen sich selbstorganisierende

[22] Trotz des scheinbaren begrifflichen Widerspruchs ist ein nichtrechtsfähiger Verein eine teilrechtsfähige Personengemeinschaft.

[23] Und somit nicht rechtsfähige.

Empfehlungssysteme nicht als Vereine klassifizieren. So kann zwar vom Bestehen eines gemeinsamen Zwecks – der Erstellung von Empfehlungen – ausgegangen werden. Problematischer ist aber schon das Bestehen einer Satzung. Eine solche bedarf zwar keiner Form und kann auch konkludent beschlossen werden (vgl. Heinrichs in [Pala06, Rn. 6 zu § 54]). Sie kann aber wohl nicht schon in der Einigung der Teilnehmer gesehen werden, einem Peer-to-Peer-System beizutreten und für die Kommunikation ein bestimmtes Protokoll zu verwenden. Die körperschaftliche Organisation und das Führen eines Gesellschaftsnamens fehlen schließlich völlig – ein selbstorganisierendes Empfehlungssystem hat keine Organe, durch die es (rechtlich) handeln kann. Auch ist ein nichtrechtsfähiger Verein zwar auf einen wechselnden Mitgliederbestand angelegt, doch geht das Vereinsrecht nicht von der ständigen Möglichkeit des Ein- und Austretens binnen kürzester Zeitspannen aus, wie sie in Peer-to-Peer-Systemen üblich ist. Insgesamt liegt also kein nichtrechtsfähiger Verein vor.

– Die Gesellschaft bürgerlichen Rechts (GbR) ist in den §§ 705 ff. BGB geregelt. In einer solchen Gesellschaft „verpflichten sich die Gesellschafter gegenseitig, die Erreichung eines gemeinsamen Zwecks [...] zu fördern [...]“ (§ 705 BGB). Eine körperschaftliche Organisation ist hier gerade nicht gegeben (vgl. Sprau in [Pala06, Rn. 4 zu § 705]). Der gemeinsame Zweck ist zwar bei selbstorganisierenden Empfehlungssystemen gegeben, doch passt die Konzeption der § 705 ff. ansonsten nicht: Der Bestand einer GbR ist von ihren Mitgliedern abhängig – im Gegensatz zum auf wechselnden Mitgliederbestand angelegten nichtrechtsfähigen Verein (Sprau in [Pala06, Rn. 4 zu § 705]). Für Vereinigungen, denen viele Personen angehören, passen die Vorschriften über die GbR deshalb nicht [Brox05, Rn. 729]. Ein selbstorganisierendes Empfehlungssystem kann deshalb nicht als GbR eingeordnet werden.

Da die im Gesetzestext vorgesehenen Alternativen nicht in Frage kommen, liegt zunächst der Gedanke einer analogen Anwendung des § 3 Satz 1 Nr. 1, § 3 Satz 2 TDG (§ 2 Satz 1 Nr. 1, § 2 Satz 2 TMG) auf das Gesamtsystem nahe. Voraussetzung für eine Analogie ist jedoch das Vorliegen einer planwidrigen Regelungslücke[24]. Eine solche besteht nicht; vielmehr hat der Gesetzgeber die in der alten Fassung des § 3 TDG noch vorgesehenen Personenvereinigungen bewusst aus dem Katalog möglicher Diensteanbieter herausgenommen, da diese „rechtlich nicht erfasst werden können“ [Bund01b, S. 15]. Zwar vertritt Spindler [SpSG04, Rn. 3 zu § 3 TDG] dennoch die Auffassung, der Begriff sei „derart weit gefaßt, dass jeder nur denkbare Handelnde hierunter fällt“, doch kann sich dies nur auf *rechtsgeschäftliches* Handeln

[24] Eine Norm kann nur analog angewendet werden wenn sie eine ähnliche Interessenlage regelt und die Regelung auf die konkret zu entscheidende Frage übertragen werden kann, damit eine durch den Gesetzgeber nicht beabsichtigte (planwidrigen) Regelungslücke ausgefüllt werden kann und kein verfassungsrechtliches Verbots dieser Analogie vorliegt [Schw03a, S. 107 f.].

beziehen, wohingegen andere Formen des Handelns zwar denkbar, aber durch das Gesetz nicht erfasst sind. Die hinter dieser Entscheidung stehenden praktischen Erwägungen sind auch durchaus nachvollziehbar: Eine Einheit, die nicht in der Lage ist, Rechte zu erwerben oder Verbindlichkeiten einzugehen, muss auch kein Normadressat sein – auch, wenn dies rein begrifflich zu einer unbefriedigenden Situation führt, da der Diensteanbieter-Begriff des TDG bzw. des TMG und der normalsprachlich verwendete Begriff in Einzelfällen auseinanderfallen können.

Da die Teilnehmer des Systems also als Diensteanbieter ausscheiden, bleibt als einzige Möglichkeit noch die Person oder Personenvereinigung, die auf Basis eines Peer-to-Peer-Protokolls die für das Empfehlungssystem benötigte Software bereitstellt. Auch hier ergeben sich bereits aus rein praktischer Sicht Probleme:

- Auch Software wird nicht notwendigerweise von einer natürlichen oder juristischen Person bzw. einer teilrechtsfähigen Personengesellschaft bereitgestellt. Vielmehr ist ein verteilter Entwicklungsprozess, bei dem viele, nicht notwendigerweise namentlich bekannte Personen zusammenwirken, nicht unüblich.

- Es muss nicht nur ein einzelnes Softwareprodukt geben, das das Protokoll des selbstorganisierenden Empfehlungssystems implementiert.

Da diese Probleme zwar auftauchen können, aber nicht müssen, lohnt sich dennoch die Betrachtung des Falls der Identifizierbarkeit einer natürlichen oder juristischen Person (oder einer teilrechtsfähigen Personengesellschaft), die die einzige Software bereitstellt, mit der das System betrieben werden kann.

Für den Fall von Peer-to-Peer-Filesharing-Systemen hat diese Frage bereits in der Literatur Erwähnung gefunden [Koch05, S. 752]. Für den Download der Software liegt demzufolge ein Teledienst vor. Bezüglich des Systems selbst kommt [Koch05, S. 752] (wie auch Spindler/Leistner [SpLe05, S. 787]) zu dem richtigen Ergebnis, dass der Softwareanbieter kein Diensteanbieter im Sinne des § 3 S. 1 Nr. 1 TDG ist, da er die von den Nutzern gewünschten Dateien weder bereithält noch zwischenspeichert oder durchleitet, sondern mit diesen überhaupt nicht in Kontakt kommt.[25] Dieses Ergebnis lässt sich auch ohne weiteres auf Empfehlungssysteme übertragen. Zwar implementiert die Software ein Protokoll, das einem Teledienst zugrunde liegt. Doch zum Einsatz bringt es eben nicht der Anbieter der Software, sondern der konkrete Nutzer. Nutzer sind es auch, die die Inhalte, auf die es bei einem Teledienst letztlich ankommt, überwiegend zur Verfügung stellen. Auch bei selbstorganisierenden Empfehlungssystemen ist der Softwareanbieter also kein Anbieter eines Teledienstes, was den Betrieb des Systems betrifft. Anders könnte sich die Situation lediglich darstellen, wenn der Softwareanbieter zusätzliche Infrastrukturdienstleistungen erbringen

[25] Der Autor diskutiert dennoch bestehende Pflichten des Sofwareanbieters, die jedoch unabhängig von der Einordnung als Diensteanbieter im Sinne des TDG sind.

würde; dies ist jedoch für das im Rahmen dieser Arbeit zu entwickelnde System nicht vorgesehen.

Teledienst ohne Diensteanbieter? Wie sich gezeigt hat, lässt sich für das Gesamtsystem kein Diensteanbieter im Sinne des § 3 S. 1 Nr. 1 TDG (§ 2 Satz 1 Nr. 1 TMG) identifizieren. Andererseits ist das System aus funktionaler Sicht eindeutig als Teledienst (Telemedium) einzuordnen. Kann es also einen Teledienst bzw. ein Telemedium ohne einen solchen Diensteanbieter geben? Oder ist ein solcher in der Definition des Telediensts bzw. des Telemediums (§ 2 Abs. 1 TDG, § 1 Abs. 1 TMG) zwingend vorausgesetzt?

Fest steht, dass das Vorliegen eines Dienstes stets ein Anbieter-Nutzer-Verhältnis erfordert (vgl. oben S. 49; Tettenborn in [Tett99, S. 518]). Doch kann ein solches nur vorliegen, wenn der Anbieter auch „Diensteanbieter“ ist? Dafür spricht zum einen der Wortlaut: Wer sonst, wenn nicht ein Diensteanbieter, sollte einen (Tele-)dienst[26] anbieten[27]? Zum anderen knüpft keine einzige Bestimmung in TDG oder TDDSG bzw. TMG direkt an das Vorliegen eines Teledienstes (Telemediums) an; es wird ausschließlich der Diensteanbieter verpflichtet. Es besteht also keine Veranlassung, die Definition des Teledienstes bzw. des Telemediums weiter zu ziehen. Ohne nähere Diskussion der Problematik kommen auch Reber/Schorr [ReSc01, S. 676] zu diesem Ergebnis.

Für die These, dass ein Teledienst bzw. Telemedium auch ohne Diensteanbieter vorliegen kann, spricht hingegen die Systematik des Abschnitts 1 des TDG: Liegt ein Teledienst nur vor, wenn auch ein „Diensteanbieter“ vorhanden ist, so setzt die Definition des Telediensts (§ 2 Abs. 1 TDG) die des Diensteanbieters (§ 3 S. 1 Nr. 1 TDG) voraus und umgekehrt. Dies gilt ebenso im neuen Telemediengesetz (§§ 1 Abs. 1, 2 Satz 1 Nr. 1 TMG). Zwar lässt sich diese zirkuläre Abhängigkeit auflösen, doch ist fraglich, ob sie in dieser Form gewünscht war. Ebenso fraglich ist, ob der Gesetzgeber mit der Änderung der Definition des Diensteanbieters auch zugleich den Begriff des Teledienstes verändern wollte. Diese Überlegungen führen tendenziell zum Ergebnis, dass ein Teledienst durchaus auch ohne Diensteanbieter im Sinne des § 3 S. 1 Nr. 1 TDG und nach neuer Rechtslage ein Telemedium ohne Diensteanbieter im Sinne des § 2 Satz 1 Nr. 1 TMG existieren kann. Letztlich kann dies jedoch dahinstehen, da zumindest nach bisheriger Rechtslage keinerlei Konsequenzen an das bloße Vorliegen eines Telediensts geknüpft sind. Das Fehlen eines Diensteanbieters führt somit dazu, dass für einen Teledienst (ein Telemedium) schlicht keine rechtlichen Verpflichtungen mehr aus dem Teledienstegesetz (dem Telemediengesetz) erwachsen.

Dienstleistungen einzelner Knoten und *Gesamtsystem als Teledienst?* Selbst, wenn man davon ausgeht, dass das Gesamtsystem ein Teledienst bzw. Telemedium ist: Dies ändert nichts daran, dass einzelne Knoten dennoch für die Dienstleistungen, die sie im Rahmen des Sys-

[26] Das Argument gilt auch für Telemedien, die nach der Definition des § 1 Abs. 1 Satz 1 TMG ebenfalls „Dienste“ sind.

[27] Dieses Kriterium ist jedoch kein sehr gewichtiges Argument, denn umgekehrt müssen Diensteanbieter gemäß § 3 S. 1 Nr. 1 TDG (§ 2 Satz 1 Nr. 1 TMG) auch nicht notwendigerweise Dienste anbieten; es genügt, wenn sie „den Zugang zur Nutzung vermitteln“.

tems erbringen, Diensteanbieter sind. Ein ähnlich gelagerter Fall liegt beispielsweise bei Online-Auktionshäusern vor: Auch hier gibt es einen übergeordneten Teledienst, nämlich das Angebot des Plattformbetreibers, und zahlreiche einzelne Teledienste bzw. Telemedien, nämlich die Angebote einzelner Warenanbieter. Die Rechtsprechung hat beide richtig als Teledienste eingeordnet (für den Plattformbetreiber [OLGBR04], für den einzelnen Warenanbieter [OLG 06]). Jedenfalls ändert also die Einordnung des Gesamtsystems als Teledienst bzw. Telemedium nichts daran, dass auch Teildienste Teledienste bzw. Telemedien sein können. In der Sache besteht zwar insofern ein Unterschied, als der Betreiber einer zentralen Plattform selbst Diensteanbieter ist und einen Dienst erbringt, der von den einzelnen Angeboten zunächst unabhängig ist; aber auch in einem selbstorganisierenden, verteilten Empfehlungssystem bietet das Systemverhalten einen Mehrwert im Vergleich zur Summe der einzelnen Knoten. Somit ist aus Nutzersicht kein grundlegender Unterschied zwischen beiden Fällen ersichtlich.

Als Ergebnis kann insgesamt festgehalten werden, dass die einzelnen Knoten des Systems als Diensteanbieter einzuordnen sind. Das Gesamtsystem kann, trotz bestehender Zweifel – aber ohne rechtliche Konsequenzen – als Teledienst (Telemedium) eingeordnet werden, für das sich jedoch kein Diensteanbieter identifizieren lässt.

Die Ergebnisse des vorliegenden Abschnitts sind in [Sorg07c] veröffentlicht.

3.6 Datenschutz in verteilten, selbstorganisierenden Systemen aus rechtlicher Sicht

Mit dieser Einordnung selbstorganisierender Empfehlungssysteme ist die Grundlage geschaffen, die Konsequenzen zu untersuchen, die sich für teilnehmende Knoten ergeben. Zum einen gilt das Verbot mit Erlaubnisvorbehalt des § 4 Abs. 1 BDSG und des § 3 Abs. 1 TDDSG (§ 12 Abs. 1 TMG). Entsprechende Erlaubnisnormen finden sich für die Inhaltsebene im BDSG, für die Interaktionsebene im TDDSG (TMG). Zum anderen werden aber auch konkrete Anforderungen an die Gestaltung datenverarbeitender Systeme gestellt; man spricht hier von *Systemdatenschutz*.

3.6.1 Abgrenzung Inhalts- und Interaktionsebene

In einem ersten Schritt gilt es nun herauszuarbeiten, wo sich bei einem selbstorganisierenden Empfehlungssystem die Grenze zwischen Inhalts- und Interaktionsebene im Sinne des Schichtenmodells (Abschnitt 3.2.2.1, S. 35) ziehen lässt.

Das TDDSG galt für den „Schutz personenbezogener Daten der Nutzer von Telediensten„ (§ 1 Abs. 1 Satz 1 TDDSG); seine Regelungen wurden überwiegend wörtlich in das TMG übernommen. Ausschlaggebend ist also – mit der herrschenden Meinung (h.M.) von der funktionalen Abgrenzung der Schichten des Datenschutz-Schichtenmodells ausgehend

(vgl. oben S. 35) – die Abgrenzung, welche Daten in der Eigenschaft als Nutzer eines Teledienstes bzw. Telemediums anfallen.

Zweifelsfrei der Interaktionsebene zuordnen lassen sich somit Daten über den Ablauf des Zugriffs auf ein System, beispielsweise Uhrzeit, IP-Adressen beteiligter Knoten, Pseudonyme der jeweiligen Nutzer bei Abruf einer Empfehlung oder Absenden eines Empfehlungsdokuments.

Sofern in Dokumenten, die über das System übermittelt werden, von Menschen erstellte personenbezogene Daten enthalten sind – beispielsweise, wenn das System die Bewertung von Systemteilnehmern erlaubt –, so handelt es sich hierbei ebenso eindeutig um Inhaltsdaten (anderer Auffassung aber Imhof [Imho00, S. 113 ff.], der die Existenz von Inhaltsdaten bei Telediensten generell ablehnt).

Es fallen jedoch auch Daten an, deren Einordnung sich schwieriger gestaltet: Daten über Präferenzen von Nutzern oder Daten, aus denen sich diese Präferenzen ableiten lassen, sind Inhalte, die zwar durch das Telemedium erstellt werden; doch sind sie nicht für Telemedien spezifisch, sondern auch ohne das Vorliegen von Telekommunikation vorstellbar. Diese Art der Abgrenzung würde jedoch zu weit führen; grundsätzlich könnte fast jede einem Telemedium zugrunde liegende Datenbasis auch ohne Telekommunikation verwendet werden. Vielmehr ist (mit Schulz in [Roßn99, 3. Teil, Rn. 43 zu § 1 TDDSG]) darauf abzustellen, ob Daten zur Durchführung eines Telemediums erhoben, verarbeitet oder genutzt werden. Lediglich bei einer reinen Übermittlung durch Telemedien (beispielsweise Daten, die für den Abschluss eines Kaufvertrags mit einem Versandhändler nötig sind) ist für die Inhaltsdaten der Anwendungsbereich des BDSG eröffnet. Der Anwendungsbereich der datenschutzrechtlichen Regelungen des TMG ist somit recht weit gefasst; dies lag aber auch in der Intention des Gesetzgebers: Die Bundesregierung lehnte im Gesetzgebungsverfahren einen Änderungsvorschlag des Bundesrats [Bund97, S. 53] zu § 1 TDDSG, der den Anwendungsbereich des Gesetzes eingeschränkt hätte, ab und beharrte auf einer umfassenden Gültigkeit des TDDSG ([Bund97, S. 70], vgl. auch Schulz in [Roßn99, 3. Teil, Rn. 12 zu § 1 TDDSG]) – die von der Bundesregierung vorgeschlagene Fassung trat schließlich auch in Kraft. Die später erfolgte Abänderung diente lediglich der Klarstellung und Verdeutlichung des Geltungsbereichs [Bund01b, S. 14]. Mit Übernahme in das TMG wurde der Anwendungsbereich der datenschutzrechtlichen Regelungen lediglich gegenüber dem TKG abgegrenzt; gegenüber dem BDSG wurde jedoch keine Abgrenzung vorgenommen. Insofern ist also von einem unveränderten Anwendungsbereich auszugehen. Im konkreten Fall von Empfehlungssystemen unterfallen somit alle erhobenen, verarbeiteten und genutzten Daten dem TMG. Ausgenommen sind nur die bereits erwähnten, von Menschen verfassten Dokumente, die durch den Teledienst nicht verarbeitet, sondern lediglich übermittelt werden. Es gilt jedoch anzumerken, dass die Einordnung von Inhaltsdaten in der Literatur

umstritten ist; eine Diskussion zur Anwendbarkeit von TDDSG und BDSG auf Inhaltsdaten findet sich in Schmitz [SpSG04, Rn. 20 ff. zu § 6 TDDSG].

3.6.2 Erlaubnistatbestände auf Interaktionsebene

Auf der Interaktionsebene – wie auch auf der Inhaltsebene – gilt das Prinzip des Verbots mit Erlaubnisvorbehalt, das hier in § 12 Abs. 1 TMG normiert ist. Zentrale Erlaubnisnormen sind – entgegen ihrem Wortlaut – die §§ 14, 15 TMG. § 14 befasst sich mit Bestandsdaten, § 15 mit Nutzungsdaten. Für Telemedien sind diese Erlaubnistatbestände auf der Interaktionsebene auch abschließend; solche des BDSG dürfen nicht ergänzend herangezogen werden (vgl. dazu bereits für Teledienste die Gesetzesbegründung des TDDSG [Bund01b, S. 14, 29] sowie Schmitz in [SpSG04, Rn. 1 zu § 5 TDDSG; Rn. 3 zu § 6 TTDSG]; der Gesetzgeber hat mittlerweile klargestellt, dass nur Erlaubnistatbestände des TMG oder solche, die sich ausdrücklich auf Telemedien beziehen, in Frage kommen (§ 12 Abs. 1 TMG)).

3.6.2.1 Bestandsdaten

Bestandsdaten sind Daten, die für die „Begründung, inhaltliche Ausgestaltung oder Änderung eines Vertragsverhältnisses mit ihm [dem Nutzer] über die Nutzung von Telediensten erforderlich sind“ (§ 14 Satz 1 TMG). Beispiele für Bestandsdaten sind die Rechnungsadresse, Authentifizierungsdaten oder die Kontakt-E-Mail-Adresse des Nutzers. Die Erhebung, Verarbeitung und Nutzung von Bestandsdaten wird durch 14 Satz 1 TMG auch ohne Einwilligung erlaubt. Die Formulierung der Vorschrift, nach der personenbezogene Daten *nur* erhoben, verarbeitet oder genutzt werden dürfen, wenn es sich um Bestandsdaten handelt, ist als gesetzgeberisches Versehen zu werten. Bei wortgetreuer Auslegung verletzt sie nicht nur die Systematik des TMG – Verbot der Verarbeitung personenbezogener Daten in § 12 Abs. 1 TMG und Durchbrechen dieses Verbots durch einzelne Erlaubnisnormen –, sondern sie setzt sich auch in Widerspruch zu § 15 Abs. 1 Satz 1 TMG, der wiederum *nur* die Erhebung, Verarbeitung und Nutzung von Nutzungsdaten zulässt. Mit der entsprechenden Änderung der §§ 5, 6 TDDSG[28] durch den Bundestagsausschuss für Wirtschaft und Technologie wurde bei der Novellierung des Gesetzes jedoch nicht das Ziel verfolgt, den Erlaubnistatbestand inhaltlich zu ändern; vielmehr sollte sie lediglich die Einführung eines Bußgeldtatbestands erleichtern [Bund01a, S. 32]. Insofern kann weiterhin von einer Erlaubnisnorm ausgegangen werden. Für ein selbstorganisierendes Empfehlungssystem problematisch könnte indes der Inhalt dieser Erlaubnisnorm sein: Der Bestandsdatenbegriff des TDDSG erfordert das Vorliegen eines Vertragsverhältnisses „über die Nutzung von Telediensten“.

Schleipfer [Schl04, S. 731 f.] hält die Formulierung des § 5 Satz 1 TDDSG (neu: § 14 Abs. 1 TMG) für einen Irrtum des Gesetzgebers und weist darauf hin, dass in vielen Fällen Be-

[28] Diese wurden wortgleich in das TMG übernommen.

standsdaten benötigt würden, jedoch lediglich ein Gefälligkeitsverhältnis vorliege. Dieser Auffassung kann aber nicht zugestimmt werden: Der Begriff der Bestandsdaten ist im TMG vielmehr enger gefasst als im allgemeinen Sprachgebrauch[29], in dem er alle personenbezogenen Daten erfasst, die einem Betroffenen auf Dauer zugeordnet sind (vgl. [Scha02, Rn. 374]). Anzeichen dafür, dass der Gesetzgeber mit § 14 Satz 1 TMG auch Gefälligkeitsverhältnisse erfassen wollte, bestehen nicht.

Ein solches Gefälligkeitsverhältnis ist auch im Fall des geplanten selbstorganisierenden Empfehlungssystems gegeben: Ein Vertrag kommt mit Wirksamwerden zweier übereinstimmender, aufeinander bezogener Willenserklärungen zustande [Brox05, Rn. 76]. Voraussetzung für eine fehlerfreie Willenserklärung ist der Rechtsbindungswille des Erklärenden. Bei der Nutzung des geplanten Systems wird jedoch in der Regel kein solcher Rechtsbindungswille vorliegen. Zwar sollen bestimmte Verhaltensweisen der Nutzer – beispielsweise das Leisten eines Beitrags durch Bereitstellung von Ressourcen oder Verfassen von Empfehlungsdokumenten – gefördert werden, doch eine rechtliche Verpflichtung soll damit nicht verbunden sein. Mit dem Fehlen des Rechtsbindungswillens kommen keine Willenserklärungen und folglich auch keine Vertragsverhältnisses zustande; mithin liegen also auch keine Bestandsdaten im Sinne des § 14 TMG vor, so dass diese Erlaubnisnorm für das geplante selbstorganisierende Empfehlungssystem nicht weiterhilft.

Dieses Ergebnis lässt sich allerdings nicht beliebig verallgemeinern. So sind Vertragsverhältnisse selbstverständlich auch in Peer-to-Peer-Systemen möglich. Es könnte durchaus eine Gegenleistung für die Berechnung einer Empfehlung verlangt und ein Vertrag darüber geschlossen werden. Selbst in diesem Fall würde die gesetzliche Erlaubnis des § 14 TMG aber keinen Ausweg bieten, denn in einem Peer-to-Peer-System interagieren oft wechselnde Knoten miteinander; die Erlaubnis für einzelne Knoten, jeweils Daten über ihren Interaktions- und Vertragspartner zu erheben, zu verarbeiten oder zu nutzen, hilft also zwar bei der Abwicklung eines Vertragsverhältnisses, nicht aber bei der Erbringung der Funktionalität des Systems weiter.

Diese Beschränkung des § 14 TMG bedeutet jedoch andererseits alleine noch keine Einschränkung der Funktionalität des geplanten Empfehlungssystems. Neben der Möglichkeit der Einwilligung besteht noch der Erlaubnistatbestand des § 15 TMG, der im Folgenden diskutiert wird.

3.6.2.2 Nutzungsdaten

§ 15 Abs. 1 Satz 1 TMG definiert Nutzungsdaten als Daten, die erforderlich sind, um „die Inanspruchnahme von Telediensten zu ermöglichen und abzurechnen“. Ihre Erhebung, Verarbeitung und Nutzung wird durch die Norm erlaubt. Der Begriff der Erforderlichkeit ist

[29] Aus dieser Sichtweise heraus erübrigt sich auch Schleipfers Einwand in [Schl04, S. 731 f.]

jedoch eng auszulegen (Schmitz in [SpSG04, Rn. 5 zu § 5 TDDSG][30]). Aus dem Vorbehalt der Erforderlichkeit folgt auch eine Pflicht zur Löschung, sobald die Erforderlichkeit wegfällt (Schmitz ebenda).

Für das Szenario des selbstorganisierenden Empfehlungssystems bedeutet dies, dass alle Daten eines Nutzers, die benötigt werden, um ihm den Dienst zu erbringen, für diesen Zweck auch erhoben, verarbeitet und genutzt werden dürfen. Aus der engen Auslegung des Begriffs der Erforderlichkeit folgt nicht, dass nur eine Version des Empfehlungssystem implementiert werden darf, die besonders wenige Daten benötigt; um die Qualität von Empfehlungen zu verbessern oder schlicht als Folge der Wahl eines bestimmten Ansatzes (wie z.B. Demographie-basiertes Filtern) dürfen auch mehr Daten verarbeitet werden, als dies grundsätzlich für irgendein Empfehlungssystem nötig wäre. Eine Verpflichtung zur datensparsamen Realisierung des Dienstes kann sich allenfalls aus § 3a BDSG ergeben, der jedoch das Angebot „datenintensiver“ Dienste ebenfalls nicht generell verbietet. Ein solches Verbot wäre auch als Eingriff in das informationelle Selbstbestimmungsrecht des Nutzers verfassungsrechtlich problematisch (vgl. dazu auch Schmitz in [SpSG04, Rn. 12 zu § 6 TDDSG]). Lediglich Daten, die auch für den gewählten Ansatz nicht nötig sind, unterfallen dem Erlaubnistatbestand des § 6 Abs. 1 Satz 1 TDDSG also nicht. Der Begriff der Nutzungsdaten ist im übrigen auch nicht disjunkt zu dem der Bestandsdaten; auch dauerhaft einem Nutzer zugeordnete Daten wie Authentifizierungsdaten (die in § 15 Abs. 1 Satz 2 Nr. 1 TMG gar als Regelbeispiel genannt werden) können Nutzungsdaten sein.

Es stellt sich jedoch die Frage, wie mit Daten umzugehen ist, die nicht oder nicht ausschließlich dem Nutzer dienen, dem sie als Person zugeordnet sind. Wertet das System beispielsweise aus, welche Objekte ein Nutzer zusammen mit anderen Objekten genutzt hat, so ist diese Information nur wertvoll, wenn sie auch für andere Nutzer – die gerade keine Empfehlung wünschen – ausgewertet wird. Der Wortlaut des § 15 Abs. 1 Satz 1 TMG scheint auch diese Nutzung personenbezogener Daten zuzulassen: Die Vorschrift stellt auf die Erforderlichkeit für die Ermöglichung der Inanspruchnahme von Telediensten ab, ohne dies jedoch darauf einzuschränken, dass diese Inanspruchnahme durch den Betroffenen erfolgt. Eine solche Auslegung würde indes dem Schutzzweck des TDDSG nicht gerecht werden. Die Transparenz über im Umlauf befindliche Daten, Voraussetzung für die Verwirklichung des informationellen Selbstbestimmungsrechts – wie es durch das TMG konkretisiert wird –, könnte gerade nicht mehr gewährleistet werden, würde die Erlaubnisnorm des § 15 TMG so weit gehen: Der Nutzer eines Teledienstes rechnet üblicherweise nicht damit, dass seine Daten auch dazu verwendet werden, anderen einen Dienst anzubieten. Dies mag sich im Fall von Empfehlungssystemen zwar anders darstellen; das TDDSG gilt jedoch für alle Teledienste. Da § 15 TMG nicht danach differenziert, ob der Nutzer mit der weiteren Ver-

[30] Auch der Wortlaut dieser Norm wurde mit Übernahme in das TMG nicht geändert.

arbeitung und Nutzung seiner Daten rechnen konnte, besteht auch für diesen Spezialfall keine Erlaubnis.

Um die Verarbeitung solcher Daten dennoch zu ermöglichen, bestehen zwei Möglichkeiten:

- Durch Anonymisierung oder Pseudonymisierung könnte der Personenbezug entfernt werden.
- Eine Einwilligung des Nutzers könnte eingeholt werden.

Diese Möglichkeiten werden in den folgenden Abschnitten diskutiert.

3.6.3 Anonymisierung und Pseudonymisierung

Wie bereits erwähnt, dient das Datenschutzrecht dem Schutz personenbezogener Daten. Sowohl das BDSG (s. § 1 Abs. 1) als auch das TMG (§ 12 Abs. 1) knüpfen an diesen Begriff an, der in § 3 Abs. 1 BDSG definiert ist als „Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (Betroffener)“. Bestimmt ist eine Person dabei, wenn aus den Daten unmittelbar auf die Identität des Betroffenen geschlossen werden kann; bestimmbar ist sie, wenn dieser Rückschluss mit Hilfe von Zusatzinformationen möglich ist (vgl. Tinnefeld in [Roßn03, Abschnitt 4.1, Rn. 20]).

Anonyme Daten zeichnen sich gerade dadurch aus, dass sie zwar Daten über eine Person enthalten, sie aber keiner Person – nicht einmal einer mehr als äußerst kurzfristig geltenden Identität – zugeordnet werden können. Sie sind daher nicht als personenbezogene Daten im Sinne des Datenschutzrechts anzusehen. Es wird nicht gefordert, dass die Herstellung des Personenbezugs beweisbar unmöglich ist; ist ein unverhältnismäßig großer „Aufwand an Zeit, Kosten und Arbeitskraft“ nötig, um den Personenbezug herzustellen, so reicht dies aus (§ 3 Abs. 6 BDSG; für eine ausführliche Diskussion der Personenbeziehbarkeit anonymer Daten siehe [RoSc00, S. 726 f.]). Anonyme Daten können beliebig erhoben, verarbeitet und genutzt werden.

Der ausschließlichen Verwendung anonymer Daten in Empfehlungssystemen stehen indes praktische Gründe entgegen: Um Manipulationen vorbeugen und Reputationsmechanismen einsetzen zu können, sind langfristig gültige Identitäten unumgänglich. Diese Überlegung führt zur Verwendung von Pseudonymen (§ 3 Abs. 7 BDSG) – doch wie sind diese einzuordnen? Das BDSG selbst trifft hierzu keine Aussage; lediglich als Regelung zum Systemdatenschutz fordert § 3a BDSG den Einsatz der Pseudonymisierung, soweit dies mit vertretbarem Aufwand möglich ist. Auch das TDDSG trifft keine eindeutige Aussage bezüglich des Personenbezugs pseudonymisierter Daten.

In der Literatur wird die „Relativität des Personenbezugs“ (Tinnefeld in [Roßn03, Abschnitt 4.1, Rn. 30]) hervorgehoben: Kennt der Datenverwender die Zuordnungsregel von

Pseudonym zu „echter“ Identität des Pseudonymträgers, so sind die Daten als personenbezogene Daten zu werten.

Diskussionswürdig ist jedoch der Umgang mit dem Risiko der Aufdeckung eines Pseudonyms. Sind einem Pseudonym zu viele Daten zugeordnet, so ist denkbar, dass dieses Profil nur noch auf eine einzige Person zutrifft – ein Risiko, das beispielsweise in [RKMG+01] diskutiert wird. Hier lässt sich jedoch keine scharfe Grenze ziehen – mit [RoSc00, S. 726 f.] ist davon auszugehen, dass der praktische Ausschluss des Personenbezugs für den Verwender der Daten ausreicht. Als Mindestvoraussetzung für diesen praktischen Ausschluss muss gelten, dass an der Herstellung des Personenbezugs für den Verwender kein wirtschaftliches Interesse besteht, die entstehenden Kosten den wirtschaftlichen Nutzen personenbezogener Daten also deutlich überwiegen. Darüber hinaus ist jedoch auch der Tatsache Rechnung zu tragen, dass nicht-wirtschaftliche Interessen zu einer Aufdeckung von Pseudonymen führen könnten. Das gilt auch dann, wenn der Personenbezug beispielsweise durch aufwendige Rechenoperationen – wie einem Brute-Force-Angriff auf eine Hash-Funktion – möglich ist. Eine feste Regel, wann von einem praktischen Ausschluss des Personenbezugs auszugehen ist, kann somit nicht angegeben werden; vielmehr ist eine Einzelfallbetrachtung vorzunehmen.

3.6.4 Einwilligung

Soll auf eine hinreichend sichere Pseudonymisierung verzichtet werden, so lässt sich die Datenverwendung weiterhin durch eine Einwilligung des Nutzers rechtfertigen. Im Rahmen der Erbringung des Teledienstes ist für den Diensteanbieter wünschenswert, diese Einwilligung ohne Medienbruch, d.h. ebenfalls auf dem Weg der Telekommunikation, einholen zu können.

Grundsätze zur datenschutzrechtlichen Einwilligung bei Telediensten sind in § 12 Abs. 3–4, § 13 Abs. 2–3 TMG geregelt. Auch die elektronische Einwilligung wird für zulässig erklärt (§ 13 Abs. 2 TMG[31]), jedoch an bestimmte Voraussetzungen geknüpft. Demnach muss sichergestellt sein, dass

- „der Nutzer seine Einwilligung bewusst und eindeutig erteilt hat“ (§ 13 Abs. 2 Nr. 1). Dies lässt sich durch Gestaltung der Nutzeroberfläche der Software, die den Dienst implementiert, erreichen. Dem Nutzer muss deutlich werden, in was er einwilligt; im „Kleingedruckten“ enthaltene Klauseln zur Datenverwendung reichen nicht aus.

- „die Einwilligung protokolliert wird“ (§ 13 Abs. 2 Nr. 2). Von Bedeutung ist dies insbesondere, wenn der Inhalt der Einwilligung sich über die Zeit ändert. Aus technischer Sicht lässt sich die Protokollierung realisieren, ohne dass für jeden einzelnen

[31] Die Norm wurde mit der Überleitung ins TMG an die entsprechende Norm (§ 94) des TKG angepasst.

Nutzer der vollständige Wortlaut aufgezeichnet wird. Einmaliges Vorhalten des Einwilligungswortlauts reicht aus, wenn für jeden Nutzer ein Zeiger auf den vollständigen Wortlaut protokolliert wird – beispielsweise könnte dies ein kryptographischer Hash über die abgegebene Erklärung oder ein Schlüssel des entsprechenden Texts in einer Datenbanktabelle sein. Die Protokollierung ist im Fall eines selbstorganisierenden, verteilten Systems nicht trivial zu realisieren: Es ist zwar davon auszugehen, dass ein Diensteanbieter die Speicherung des Protokolls delegieren darf und somit nicht jeder Knoten, mit dem ein Teilnehmer in Kontakt tritt, die Einwilligung speichern muss; dennoch ist aber die Speicherung auf lediglich einem einzelnen Knoten nicht ausreichend, soll die Verfügbarkeit der protokollierten Daten gewährleistet werden. Die hierbei auftretenden Schwierigkeiten sind jedoch lösbar, denn eine robuste Speicherung von Daten gehört ohnehin zu den Anforderungen, die an viele Peer-to-Peer-Systeme gestellt werden.

- „der Inhalt der Einwilligung jederzeit vom Nutzer abgerufen werden kann." (§ 13 Abs. 2 Nr. 3). Wenn ohnehin eine Protokollierung implementiert ist, ist diese zusätzliche Anforderung unproblematisch zu erfüllen.

- „der Nutzer die Einwilligung jederzeit mit Wirkung für die Zukunft widerrufen kann." (§ 13 Abs. 2 Nr. 4). Hierbei handelt es sich nicht um die Rücknahme einer Einwilligung – diese Möglichkeit einer solchen Rücknahme, die sich auch auf die Vergangenheit erstreckte, wurde durch den Gesetzgeber in der wortgleichen Regelung des § 94 TKG bewusst gestrichen (Büttgen in [GPSS06, Rn. 12 zu § 94 TKG]). Der Widerruf einer Einwilligung ist technisch zunächst unproblematisch; solange lesend auf die Einwilligungserklärung zugegriffen werden kann, ist grundsätzlich auch die Möglichkeit einer Löschung gegeben. Lediglich eine Möglichkeit der Authentifizierung müsste dazu geschaffen werden. Wenn allerdings ein robustes System gespeicherte Daten auf mehrere Knoten repliziert, können praktische Probleme auftreten; es muss sichergestellt werden, dass die Aufforderung zur Löschung auch alle diese Knoten erreicht und von ihnen befolgt wird.

Insgesamt kann die Einwilligung also auch in einem selbstorganisierenden, verteilten System – trotz auftretender Schwierigkeiten bei der Protokollierung – entsprechend den Anforderungen des TMG erklärt werden.

3.6.5 Erlaubnistatbestände auf Inhaltsebene

Wie bereits erwähnt, ist der Anwendungsbereich des BDSG im Fall selbstorganisierender Empfehlungssysteme sehr eingeschränkt. Hauptsächlicher Anwendungsfall sind von den Nutzern abgegebene textuelle Bewertungen. Als Erlaubnistatbestand kommt auf Inhaltsebene insbesondere § 28 BDSG in Frage. Absatz 1 Nr. 1 der Vorschrift erlaubt „Erheben,

Speichern, Verändern oder Übermitteln personenbezogener Daten oder ihre Nutzung als Mittel für die Erfüllung eigener Geschäftszwecke [...] wenn es der Zweckbestimmung eines Vertragsverhältnisses oder vertragsähnlichen Vertrauensverhältnisses mit dem Betroffenen dient".

Zwar liegt – wie bereits diskutiert – im Fall des Empfehlungssystems kein Vertragsverhältnis vor. Von einem vertragsähnlichen Vertrauensverhältnis ist aber auszugehen: Zwar liegt kein Rechtsbindungswille bezüglich der Verbreitung von Empfehlungsdokumenten vor, doch erledigt der Diensteanbieter eine Aufgabe auf Wunsch seines Nutzers, der ein solches Dokument verbreiten will. Die beiden beteiligten Parteien setzen dabei Vertrauen ineinander: Es liegt im Interesse des Nutzers, dass sein Dokument Verbreitung findet[32]. Umgekehrt erwartet der Diensteanbieter vom Nutzer, dass dessen Empfehlungsdokument einen Mehrwert für das Gesamtsystem bedeutet.

Die Erlaubnis des § 28 Abs. 1 Nr. 1 BDSG ist für selbstorganisierende Empfehlungssysteme also anwendbar.

3.6.6 Erstellung von Nutzungsprofilen

Um Empfehlungen zu generieren, könnten – die Nutzung des Empfehlungssystems selbst oder die Nutzung anderer Dienste betreffende – sogenannte Nutzungsprofile hilfreich sein, also Daten über die Verwendung des Teledienstes bzw. Telemediums durch einen Nutzer. Grundsätzlich bestehen hier zwei Möglichkeiten der Realisierung:

- Der Nutzer selbst verwaltet sein Nutzungsprofil, und es wird lediglich lokal bei der Auswertung von Antworten des Dienstes berücksichtigt. In diesem Fall entstehen aus Sicht des Datenschutzes keine Probleme.

- Das Nutzungsprofil wird von einem oder mehreren Diensteanbietern im Netz erstellt.

Der zweite Fall ist in § 15 Abs. 3 TMG geregelt. Nach Satz 1 dieser Norm darf der Diensteanbieter unter anderem „zur bedarfsgerechten Gestaltung der Teledienste Nutzungsprofile bei Verwendung von Pseudonymen erstellen, sofern der Nutzer dem nicht widerspricht." Nutzungsprofile zu erstellen, um bessere Empfehlungen zu ermöglichen, ist nach dieser Vorschrift also gestattet, denn verbesserte Empfehlungen sind ein Fall der „bedarfsgerechten Gestaltung" des Empfehlungsdiensts. Der Nutzer ist über sein Widerspruchsrecht zu unterrichten (Satz 2 der Vorschrift). Zudem ist ein Zusammenführen von Daten über den Pseudonymträger mit dem Nutzungsprofil nicht gestattet (Satz 3).

Somit stellt sich aber die Frage, ob § 15 Abs. 3 die Geltung des TMG nicht über personenbezogene Daten hinaus erweitert, indem dem Diensteanbieter auch für den Fall pseudonymisierter Daten Pflichten auferlegt werden – insbesondere unter Berücksichtigung des

[32] Die Verbreitung bringt ihm zwar keinen unmittelbaren Vorteil, doch würde er diesen Beitrag nicht als in seinem Interesse liegend erachten, würde er ihn nicht leisten; seinerseits profitiert der Nutzer von anderen, ähnlich handelnden Systemteilnehmern.

Verbots aus Satz 3 der Vorschrift. Dies ist jedoch nicht der Fall; es bleibt bei der sich aus § 12 ergebenden Geltung nur für personenbezogene Daten. Auch bei der Verwendung von Pseudonymen kann aber ein solcher Personenbezug bestehen. Schmitz [SpSG04, Rn. 25 zu § 6 TDDSG] weist sogar – zumindest für den Regelfall völlig zu recht – darauf hin, dass ein möglicher Personenbezug sogar erforderlich ist, um bei mehreren Nutzungsvorgängen anfallende Daten überhaupt erst einem Pseudonym zuordnen zu können. § 15 Abs. 3 TMG stellt also in der Tat eine zusätzliche Erlaubnisnorm dar, die aber nur greift, wenn der Anbieter die Pseudonyme grundsätzlich mit einer Person in Verbindung bringen kann. Das Verbot dieser Zusammenführung ändert nichts am Personenbezug, der sich nämlich nicht aus dem rechtlich Erlaubten, sondern allein nach dem technisch Praktikablen ergibt.

3.6.7 Datenschutzrechtliche Pflichten und Systemdatenschutz für Teledienste

Neben der Regelung von Erlaubnistatbeständen für die Verwendung personenbezogener Daten beinhaltet das TMG auch Aspekte des Systemdatenschutzes. Sie sind überwiegend in der Kontrollebene des in Abschnitt 3.1.1.1 eingeführten Modells einzuordnen. Konkret ergeben sich – unter der Annahme eines kostenfreien Systems, bei dem keine Abrechnungsdaten benötigt werden[33] – als Pflichten

- die Unterrichtung des Nutzers über Art, Umfang und Zweck der Erhebung und Verwendung seiner personenbezogenen Daten (§ 13 Abs. 1 TMG).
- die Ermöglichung der Beendigung der Nutzung des Dienstes (§ 13 Abs. 4 Nr. 1 TMG).
- die Sicherstellung der sofortigen Löschung oder Sperrung der angefallenen Nutzungsdaten nach Zugriff (§ 13 Abs. 4 Nr. 2 TMG).
- den Schutz der Benutzung des Dienstes vor der Kenntnisnahme Dritter (§ 13 Abs. 4 Nr. 3 TMG).
- die Ermöglichung der getrennten Verwendung von Daten über die Nutzung verschiedener Telemedien (§ 13 Abs. 4 Nr. 4 TMG).
- die lediglich pseudonyme Erstellung von Nutzungsprofilen, die nicht mit Daten über die Pseudonymträger zusammengeführt werden dürfen (§ 13 Abs. 4 Nr. 6 TMG).
- die Anzeige der Weitervermittlung zu einem anderen Diensteanbieter (§ 13 Abs. 5 TMG).
- die Ermöglichung anonymer oder pseudonymer Nutzung des Dienstes (§ 13 Abs. 6 TMG).

[33] Werden Abrechnungsdaten benötigt, so ergeben sich für den Umgang mit diesen Daten zusätzliche Pflichten.

- die Pflicht zur Auskunftserteilung über gespeicherte, personenbezogene oder pseudonyme Daten (§ 13 Abs. 7 TMG).
- die Pflicht zur Anbieterkennzeichnung (§ 5 TMG).

Diese Pflichten werden in den folgenden Abschnitten diskutiert. Sie richten sich ausschließlich an den Diensteanbieter – auch wenn ihre Umsetzung überwiegend nur in der Entwurfsphase des Systems (also zu einem Zeitpunkt, zu dem es noch gar keinen Diensteanbieter gibt) bzw. bei der Implementierung der Software, die die Systemfunktionalität umsetzt, möglich ist.

3.6.7.1 Unterrichtungspflicht

Gemäß § 13 Abs. 1 S. 1 TMG ist der Nutzer durch den Diensteanbieter „zu Beginn des Nutzungsvorgangs über Art, Umfang und Zwecke der Erhebung und Verwendung personenbezogener Daten [...] zu unterrichten, sofern eine solche Unterrichtung nicht bereits erfolgt ist." Diese Verpflichtung trifft zwar grundsätzlich jeden Knoten, der einen Dienst für den Nutzer erbringt, doch könnte der zweite Halbsatz eine Lösung bieten: Wird die Unterrichtung bereits durch die Software vorgenommen, die das Protokoll implementiert, so ist eine erneute Unterrichtung durch den einzelnen Diensteanbieter grundsätzlich nicht mehr nötig. Zwei Probleme treten bei diesem Lösungsansatz jedoch auf:

- Es ist nicht sichergestellt, dass ein Nutzer tatsächlich die vom Diensteanbieter erwartete Software einsetzt. Eine andere Software führt die erwünschte Unterrichtung möglicherweise aber nicht durch. Das vergleichbare Problem der Darstellung der Anbieterkennzeichnung nach § 6 TDG (neu: § 5 TMG) durch Web-Browser wurde in Literatur und Rechtsprechung bereits betrachtet. So fordert Woitke [Woit03, S. 873] eine Darstellung, die „insgesamt kompatibel zu den gängigen Browsern mit ihren jeweiligen Default-Einstellungen" und gängigen Änderungen dieser Einstellungen ist. Es könne nicht angenommen werden, dass Grafiken auf jedem Endsystem dargestellt werden. In [Brun04, S. 10] wird jedoch davon ausgegangen, dass seltene Ausgabegeräte (wie beispielsweise solche, die keine Grafiken darstellen) nicht berücksichtigt werden müssen. Auch Spindler (in [SpSG04, Rn. 10 zu § 6 TDG]) hält die „in den angesprochenen Verkehrskreisen übliche, durchschnittliche technische Ausrüstung" für maßgeblich. Diese Auffasung geht zwar etwas zu weit: Das Datenschutzrecht soll nicht lediglich den Durchschnittsnutzer, sondern grundsätzlich jeden schützen, dessen personenbezogene Daten verarbeitet werden. Doch kann ihr insoweit zugestimmt werden, dass zumindest sehr seltene Konfigurationen auf Nutzerseite nicht berücksichtigt werden müssen: Der Nutzer kann diese in aller Regel im Bedarfsfall mit wenig Aufwand ändern. Der Aufwand des Diensteanbieters, alle Eventualitäten zu berücksichtigen, wäre jedoch sehr hoch.

Dieses Ergebnis lässt sich auch auf die Unterrichtungspflicht des § 13 Abs. 1 S. 1 TMG übertragen; beide Vorschriften dienen in vergleichbarer Weise der Herstellung von Transparenz, im Fall des § 6 TDG (§ 5 TMG) auf Empfängerebene, im Fall des § 13 Abs. 1 S. 1 TMG auf Zweckebene.

Doch wie stellt sich die Situation dar, falls eine Software, die den Unterrichtungspflichten nicht gerecht wird, weite Verbreitung findet? Als Lösungsansatz bietet sich an, innerhalb der ausgetauschten Nachrichten die verwendete Softwareversion sowie die Information, ob und welche Unterrichtung stattgefunden hat, zu vermerken. Auf diese Weise könnte der Diensteanbieter zumindest mit hoher Wahrscheinlichkeit darauf schließen, ob der Nutzer unterrichtet wurde – dies gilt, obwohl die Übertragung der genannten Information natürlich keinen Beweis im naturwissenschaftlichen Sinn darstellt, dass die Unterrichtung tatsächlich dargestellt wurde. Die Erbringung eines solchen Beweises ist dem Diensteanbieter aber auch nicht zuzumuten. Wenn andererseits der Nutzer den Diensteanbieter bewusst täuscht, indem er vorgibt, unterrichtet worden zu sein, so muss er sich dies zurechnen lassen. Dies gilt auch, wenn das Verhalten durch die vom Nutzer verwendete Software verursacht wird.

- Es stellt sich die Frage, wie konkret die Unterrichtung sein muss. Umfasst die Unterrichtung über „Art, Umfang und Zwecke“ der Datenverarbeitung (bzw. -erhebung und -nutzung) auch die Frage, von wem genau die Daten verarbeitet werden? Dies kann eindeutig verneint werden, da der Gesetzgeber die Unterrichtungspflicht über den Ort der Datenverarbeitung bewusst aus dem Gesetz gestrichen hat: Eine solche Regelung wurde nicht mehr als praktikabel empfunden ([Bund01b, S. 28]; vgl. auch Schmitz in [SpSG04, Rn. 4 zu § 4 TDDSG]). Eine einmalige Unterrichtung durch die Software reicht also grundsätzlich aus, selbst, wenn zu diesem Zeitpunkt noch nicht klar ist, durch wen die Daten eigentlich verarbeitet werden.

Zusätzlich zu den bereits genannten Angaben hat die Unterrichtung auch einen Hinweis zu enthalten, dass die Daten auch außerhalb der Europäischen Union verarbeitet werden (auch dies eine Anforderung des § 13 Abs. 1 Satz 1 TMG); dies ist in einem Peer-to-Peer-System kaum zu vermeiden.

§ 13 Abs. 1 Satz 2 TMG enthält eine Regelung für automatisierte „Verfahren, die eine spätere Identifizierung des Nutzers ermöglichen und eine Erhebung oder Verwendung personenbezogener Daten vorbereiten“; dies betrifft beispielsweise den Fall von Cookies, kann jedoch auch für die bei einem Empfehlungssystem verwendete Software zutreffen – etwa, wenn der Nutzer eine Identität für die Benutzung des Systems auswählt und erstellt. In diesem Fall ist die Einwilligung bereits „zu Beginn dieses Verfahrens“, also nicht erst bei der tatsächlichen Erhebung oder Verarbeitung personenbezogener Daten zu unterrichten.

Nach § 13 Abs. 1 Satz 3 TMG muss der Inalt der Unterrichtung „für den Nutzer jederzeit abrufbar“ sein – wird die Unterrichtung durch die beim Nutzer installierte Software durchgeführt, so ist dieses Erfordernis unproblematisch zu erfüllen. Doch auch bei anderen Formen der Unterrichtung bestehen hier keine technischen Probleme – dies insbesondere, da mit der aktuellen Fassung des TMG keine Verpflichtung zur Protokollierung der tatsächlich durchgeführten Unterrichtung besteht (Schmitz bereits in [SpSG04, Rn. 8 zu § 4 TDDSG]).

3.6.7.2 Beendigung der Nutzung

§ 13 Abs. 4 TMG enthält Vorschriften zum Systemdatenschutz; es handelt sich um Regelungen zur technischen und organisatorischen Gestaltung von Datenverarbeitungssystemen.

§ 13 Abs. 4 Nr. 1 TMG schreibt vor, dem Benutzer die jederzeitige Beendigung der Dienstnutzung zu ermöglichen. Der Verbindungsabbruch (so der Wortlaut der entsprechenden Regelung im TDDSG) ist bei Internet-Anwendungen grundsätzlich unproblematisch (vgl. dazu auch Schmitz in [SpSG04, Rn. 32 zu § 4 TDDSG]); aus diesem Grund müssen beim Entwurf eines Empfehlungssystems hier auch keine besonderen Anforderungen beachtet werden.

3.6.7.3 Löschen von Daten nach Zugriff

Gemäß § 13 Abs. 4 Nr. 2 TMG ist sicherzustellen, dass personenbezogene Daten „über den Ablauf des Zugriffs oder der sonstigen Nutzung unmittelbar nach deren Beendigung gelöscht“ oder gesperrt werden. Für das Empfehlungssystem selbst ist diese Anforderung zunächst unproblematisch: Der Ablauf des Zugriffs auf das System ist meist nicht von Interesse; die transportierten Inhalte – beispielsweise also abgerufene Empfehlungen – sind von der Vorschrift nicht betroffen.

Dennoch kann es Fälle geben, in denen die Vorschrift zu Einschränkungen führt:

- wenn das Verhalten des Nutzers in einem anderen System beobachtet werden soll – beispielsweise der Einkauf bei Versandhändlern oder Downloads in einer Filesharing-Börse –, um Rückschlüsse auf seine Präferenzen zu gewinnen.
- wenn das Verhalten des Nutzers bei Benutzung des Empfehlungsdienstes selbst beobachtet werden soll, um beispielsweise Angriffe auf das System entdecken oder ebenfalls auf die Präferenzen des Nutzers schließen zu können.

In der entsprechenden Vorschrift des TDDSG (§ 4 Abs. 4 Nr. 2) war explizit lediglich die *Möglichkeit* zur Löschung gefordert – nicht, dass die genannten Daten in jedem Fall sofort nach Ablauf der Nutzung gelöscht werden müssen. Dies ist auch aus systematischer Sicht die zu bevorzugende Lösung: § 13 Abs. 4 ist eine Regelung des Systemdatenschutzes und dient dazu, auf die Gestaltung eines Systems einzuwirken, in dem personenbezogene Daten

erhoben und verwendet werden. Die Erhebung und Verwendung selbst ist davon zunächst unabhängig.

Auf die technische Realisierung indes wirkt sich die Änderung des Gesetzeswortlauts nicht aus, da die technischen Vorkehrungen für eine mögliche Löschung bereits nach TDDSG vorgesehen sein mussten; lediglich in der praktischen Anwendung ergibt sich die Verpflichtung, die Löschung auch tatsächlich durchzuführen.

3.6.7.4 Schutz vor Kenntnisnahme Dritter

§ 13 Abs. 4 Nr. 3 TMG schreibt vor, dass Nutzer Teledienste „gegen Kenntnisnahme Dritter geschützt“ benutzen können müssen. Unter Beachtung des Schichtenmodells des Telekommunikationsdatenschutzes ergibt sich, dass dieser Schutz sich nicht auf die Telekommunikation selbst bezieht, sondern lediglich auf die Datenverarbeitung durch den Diensteanbieter (vgl. Schmitz in [SpSG04, Rn. 34 zu § 4 TDG]). Die Vorschrift bedeutet auch nicht, dass der Diensteanbieter sich nicht bei der Diensterbringung der Mitarbeit anderer Diensteanbieter bedienen darf: Wird ein Dienst durch mehrere Anbieter gemeinsam erbracht, so ist keiner dieser Diensteanbieter ein „Dritter“ (vgl. für den parallelen Fall des Telekommunikationsdiensteanbieter betreffenden § 88 TKG [Meng04, S. 2018]). Wer am Erbringen des Dienstes nicht beteiligt ist, soll jedoch keine Kenntnis von der Nutzung erlangen. Die Erfüllung dieser Forderung ist für ein Empfehlungssystem jedoch unproblematisch. Besteht die Nutzung in der Abgabe einer Empfehlung, so ist die Kenntnisnahme durch Dritte kaum zu vermeiden – dies ist aber auch gar nicht gefordert (vgl. Roßnagel in [Roßn03, Abschnitt 7.9, Rn. 119]). Eine solche Regelung wäre für den Schutz der Nutzer weder nötig noch zielführend.

3.6.7.5 Getrennte Verarbeitung von Daten verschiedener Teledienste

Das Gebot der „informationellen Gewaltenteilung“ (so Schmitz in [SpSG04, Rn. 35 zu § 4 TDDSG]) aus § 13 Abs. 4 Nr. 4 TMG besagt, dass es möglich sein muss, personenbezogene Daten über die Inanspruchnahme verschiedener Teledienste getrennt zu verarbeiten. Dies soll die Erstellung umfassender Nutzungsprofile verhindern (vgl. dazu die Gesetzesbegründung [Bund97, S. 24], bezogen auf die inhaltlich identische alte Fassung des § 4 Abs. 2 Nr. 4 TDDSG). Problematisch ist jedoch die Abgrenzung, wann noch ein einheitlicher Dienst und wann „verschiedene Teledienste“ (bzw. Telemedien) vorliegen (eine Problematik, auf die bereits [Büll99, S. 264] hinweist).

Nach [EFMT97, S. 2982] ist vom „konkreten Angebot einer einzelnen, in sich abgeschlossenen Informations- und Kommunikationsdienstleistung [...] und nicht vom Gesamtspektrum der Angebote eines Diensteanbieters“ auszugehen (so auch [Enge97, S. 234]) – dieser Auffassung ist zwar zuzustimmen, doch ist es nach wie vor schwierig, zu entscheiden, wann eine Dienstleistung „in sich abgeschlossen“ ist.

In der Praxis müssen zur Abgrenzung wohl mehrere Kriterien betrachtet werden.

- Die *Nutzerperspektive*: Stellt sich ein Dienst dem Nutzer als einheitlich dar, so kann er er auch mit einer zusammenhängenden Speicherung seiner Daten rechnen; er ist also weniger schutzbedürftig als bei scheinbar voneinander unabhängigen Diensten.
- Die *technische Perspektive*: Ein einheitlicher Dienst kann auch gegeben sein, wenn aus technischer Sicht ein enger Zusammenhang der Einzelkomponenten besteht und diese typischerweise gemeinsam implementiert werden. Beispiel sind Adressbuch- und E-Mail-Dienste, die in der Praxis fast immer gemeinsam angeboten werden, da von einer zusammenhängenden Nutzung ausgegangen wird und auf eine gemeinsame Authentifizierungskomponente zurückgegriffen werden kann. E-Mail-Dienste und Suchmaschinen hingegen werden zwar von mehreren Anbietern unter einer einheitlichen Bedienoberfläche angeboten, doch besteht aus technischer Sicht hier kein Zusammenhang.

In der Tendenz führt diese Einordnung dazu, dass die gemeinsame Verarbeitung von Daten über die Nutzung unterschiedlicher Dienstleistungen dann ermöglicht werden darf, wenn dazu ein technischer Grund besteht und der Benutzer mit der gemeinsamen Verarbeitung rechnen konnte. Dies erfordert jedoch stets eine Einzelfallbetrachtung.

Welche Folgen hat dies nun für Empfehlungssysteme? Es kann für ein Empfehlungssystem vorteilhaft sein, Daten über die Nutzung anderer Dienste zu verarbeiten. So könnten – wie bereits erwähnt – Präferenzen von Nutzern aus ihrem Nutzungsverhalten bei der Verwendung dieser anderen Dienste abgeleitet werden, um so die Qualität von Empfehlungen zu verbessern. Ein technischer Grund für die Integration dieser Dienste mit dem Empfehlungssystem ist also gegeben. Doch kann die gemeinsame Verarbeitung der Daten nur erfolgen, wenn dem Nutzer auch der Eindruck eines einheitlichen Dienstes vermittelt wird. Wenn dies nicht gelingt, so muss – auch im Lichte des Transparenzgedankens, wie er aus dem informationellen Selbstbestimmungsrecht abzuleiten ist – auf die mögliche Verbesserung der Empfehlungsqualität zugunsten des Datenschutzes verzichtet werden.

3.6.7.6 Trennung von Nutzungsprofilen und Daten über Pseudonymträger

§ 15 Abs. 3 TMG erlaubt die Erstellung von Nutzungsprofilen „für Zwecke der Werbung, der Marktforschung oder zur bedarfsgerechten Gestaltung der Telemedien“ (vgl. Abschnitt 3.6.6). Jedoch hat der Diensteanbieter sicherzustellen, dass diese Profile nicht mit Daten über den Pseudonymträger zusammengeführt werden können (§ 13 Abs. 4 Nr. 6 TMG). Diese Forderung ist unproblematisch zu erfüllen; im zu entwerfenden Empfehlungssystem ist die Verwaltung der Pseudonyme unabhängig von der eigentlichen Funktionalität des Systems zu realisieren.

Es gilt zu beachten, dass trotz dieser Trennung der Personenbezug der Pseudonyme – falls ein solcher besteht – nicht aufgehoben wird (§ 15 Abs. 3 TMG wäre sonst obsolet). Vielmehr können alle Daten, die erforderlich sind, um Daten über den Pseudonymträger mit seinem Nutzungsprofil zusammenzuführen, beim Diensteanbieter vorhanden sein. Dieser hat jedoch seine Datenverarbeitungsprozesse so zu gestalten, dass diese prinzipielle Möglichkeit in der Praxis nicht umgesetzt werden kann.

3.6.7.7 Anzeige der Weitervermittlung

Wird ein Nutzer zu einem anderen Diensteanbieter weitervermittelt, so ist ihm dies anzuzeigen (§ 13 Abs. 5 TMG). Ziel ist nach der Gesetzesbegründung [Bund97, S. 24] die Herstellung von Transparenz, um dem Nutzer die Inanspruchnahme seines Auskunftsrechts und eine wirksame Datenschutzkontrolle zu ermöglichen. Wenn mehrere Anbieter einen Dienst zusammen erbringen, heißt dies jedoch nicht, dass bei jeder anstehenden Teilaufgabe, die von einem anderen Anbieter erbracht wird, eine Information an den Nutzer zu erfolgen hat. Der Übergang zwischen der gemeinsamen Erbringung eines Dienstes durch mehrere Anbieter und der Weitervermittlung eines Nutzers ist zwar fließend; im Fall der Dienserbringung in einem Peer-to-Peer-System überwiegt jedoch der Charakter des gemeinsamen Angebots.

3.6.7.8 Anonyme und pseudonyme Nutzung

Nach § 13 Abs. 6 TMG hat der Diensteanbieter dem Nutzer – soweit technisch möglich und zumutbar – eine anonyme oder pseudonyme Nutzung des Teledienstes zu ermöglichen. Diese Bestimmung betrifft also die Identitätsebene des in Abschnit 3.1.1.1 vorgestellten Modells. Mit Schmitz [Schm00, S. 115] ist davon auszugehen, dass bei internetbasierten Diensten der Diensteanbieter selbst meist nicht die Möglichkeit hat, die anonyme Nutzung zu ermöglichen – mit der IP-Adresse des Nutzers liegt bereits eine Art Pseudonym vor, auch, wenn der Diensteanbieter in der Regel die Zuordnungsvorschrift zu einer natürlichen Person nicht kennt.

Die pseudonyme Nutzung hingegen ist im geplanten Empfehlungssystem jedoch realisierbar. Für die Funktionalität kann die Kenntnis einer Identität zwar hilfreich sein; diese Identität kann jedoch auch in einem Pseudonym bestehen (vgl. Abschnitt 3.6.3, S. 63). Auch diese Anforderung des TDDSG kann also umgesetzt werden.

3.6.7.9 Auskunftserteilung

§ 13 Abs. 7 Satz 1 TMG schreibt vor, dass dem Nutzer, wenn er dies verlangt, „unentgeltlich und unverzüglich Auskunft über die zu seiner Person oder zu seinem Pseudonym gespeicherten Daten zu erteilen“ ist. Jedoch ist diese Vorschrift im Fall der Verwendung von

Pseudonymen nicht unproblematisch. Es sind dann zwei mögliche Fallgruppen zu unterscheiden:

- Der Diensteanbieter hat Daten zum Nutzer unter einem Pseudonym gespeichert; dieses Pseudonym ist dem Nutzer jedoch nicht bekannt, da es durch den Anbieter generiert wurde. Der Nutzer kann sein Auskunftsbegehren in diesem Fall nur unter seiner eigentlichen Identität vorbringen. Um ihm zu entsprechen, müsste der Anbieter also Daten über den Träger des Pseudonyms mit zu diesem Pseudonym gespeicherten Daten zusammenführen. Zumindest für den in der Praxis wohl relevantesten Fall, dass es sich zumindest auch um Nutzungsprofile handelt, ist ihm das aber durch § 13 Abs. 4 Nr. 6 TMG verboten. Schmitz (in [SpSG04, Rn. 51, 52 zu § 4 TDDSG]) kommt zu dem Ergebnis, dass die resultierende Verpflichtung zur Aufdeckung des Pseudonyms das informationelle Selbstbestimmungsrecht des Nutzers verletzt und § 4 Abs. 7 TDDSG (neu: § 13 Abs. 7 TMG) insoweit verfassungswidrig und somit nichtig ist. Als mögliche – und durchaus praktikable – Lösung schlägt er vor, den Nutzer in diesem Fall über die Folgen der Auskunftserteilung zu informieren und die tatsächliche Auskunftserteilung von seiner Einwilligung abhängig zu machen. Dies kann jedoch nicht heißen, dass der Anbieter durch diese Vorschrift gezwungen werden kann, die Zusammenführung von Pseudonymen und Daten über Pseudonymträger überhaupt technisch zu ermöglichen: Damit würde sich § 13 Abs. 7 TMG in Widerspruch zum Gebot der Datenvermeidung und Datensparsamkeit des § 3a BDSG setzen. Vielmehr sind lediglich vorhandene Möglichkeiten im Falle eines Auskunftsbegehrens – und falls der Nutzer eingewilligt hat – zu nutzen.

- Dem Nutzer ist sein Pseudonym bekannt. In diesem Fall kann er unter seinem Pseudonym Auskunft verlangen; es sind dann jedoch Sicherungsmaßnahmen zu ergreifen, um zu verhindern, dass Dritte das Auskunftsbegehren vorbringen. So könnte ein gemeinsames Geheimnis zwischen Nutzer und Diensteanbieter vereinbart werden (so auch Schaar in [Scha02, Rn. 508]). Wenn der Nutzer auf das gemeinsame Geheimnis nicht mehr zugreifen kann, sollte der Anbieter im Rahmen der – eng begrenzten – technischen Möglichkeiten versuchen, alternative Authentifizierungsmechanismen anzubieten.

  Zu beachten gilt ferner, dass Auskünfte im Fall des pseudonymen Auskunftsbegehrens nicht schriftlich erteilt werden können; nach § 34 Abs. 3 BDSG, zu dem § 13 Abs. 7 TMG lex specialis ist, ist die schriftliche Auskunftserteilung der Regelfall. § 13 Abs. 7 Satz 2 TMG weicht davon ab, indem er die elektronische Auskunft „auf Verlangen des Nutzers“ erlaubt. Da die schriftliche Auskunft in aller Regel die Aufdeckung des Pseudonyms erfordert, ist ein pseudonymes Auskunftsbegehren aber als solches Verlangen zu interpretieren.

Für das geplante Empfehlungssystem ist zur Erfüllung des Auskunftsanspruchs vorzusehen, dass mit Pseudonymen gearbeitet wird, die den Nutzern bekannt sind. Eine ohnehin benötigte Authentifizierungskomponente ist auch zur Prüfung pseudonym vorgebrachter Auskunftsbegehren einzusetzen.

3.6.7.10 Informationspflichten

Die Pflicht des Anbieters, dem Nutzer gewisse Informationen zur Verfügung zu stellen, hat direkte Auswirkungen auf den Datenschutz – sowohl für den Anbieter, der ggf. personenbezogene Daten bereitstellen muss, als auch für den Nutzer, für den die Anbieterkennzeichnung Transparenz herstellen kann, wer seine personenbezogenen Daten verarbeitet. Sie betrifft somit die Kontrollebene des in Abschnitt 3.1.1.1 diskutierten Modells.

Die Informationspflichten des Anbieters sind in § 5 TMG geregelt[34]. Sie umfassen zumindest Namen und Anschrift des Anbieters (§ 5 Satz 1 Nr. 1 TMG) sowie „Angaben, die eine schnelle elektronische Kontaktaufnahme und unmittelbare Kommunikation mit [ihm] ermöglichen, einschließlich der Adresse der elektronischen Post“ (§ 5 Satz 1 Nr. 2 TMG); in den Fällen juristischer Personen, bei Vorliegen einer beruflichen oder zulassungspflichtigen Tätigkeit oder eines Eintrags in einem öffentlichen Register sowie bei Besitz einer Umsatzsteueridentifikationsnummer treten weitere Angaben hinzu. Die Anschrift muss eine ladungsfähige postalische Adresse sein (siehe nur Spindler in [SpSG04, Rn. 23 zu § 6 TDG] sowie die Gesetzesbegründung zur weitgehend übereinstimmenden Vorgängerregelung § 6 TDG [Bund01b, S. 21]). Zu den Angaben, die „eine schnelle elektronische Kontaktaufnahme und unmittelbare Kommunikation“ ermöglichen, zählt nach dieser Gesetzesbegründung [Bund01b, S. 21] auch zumindest die Telefonnummer. Auch in Literatur und Rechtsprechung wird die Pflicht zur Angabe der Telefonnummer vertreten (so – ohne nähere Begründung – beispielsweise in [Brun04, S. 10], [Stic04, S. 113]). Doch ist der Gegenmeinung zuzustimmen: Dem Gesetzestext ist keine entsprechende Verpflichtung zu entnehmen; unmittelbare Kommunikation ist grundsätzlich auch mittels Instant Messaging, E-Mail, über Web-Formulare und ähnliche Techniken möglich (vgl. dazu auch [OLGH04, S. 1046]). Mit Hinweis auf verfassungsrechtliche Bedenken aufgrund der Unbestimmtheit der Norm wird eine Verpflichtung zur Angabe der Telefonnummer auch in [Füll05, S. VI] zu Recht abgelehnt.

Doch stellt sich die Frage, ob Nutzer eines Peer-to-Peer-Systems, zumindest soweit es – wie bisher die Regel – kostenlos angeboten wird, überhaupt der Kennzeichnungspflicht des § 5 TMG unterliegen. Diese gilt nämlich lediglich für „geschäftsmäßige, in der Regel gegen Entgelt angebotene Telemedien“. Der Begriff „geschäftsmäßig“ ist umstritten, da das TMG keine Legaldefinition enthält. Mit Blick auf die Definition des § 3 Nr. 10 TKG[35] wurde in der

[34] Für „kommerzielle Kommunikationen“ gilt zusätzlich § 6 TMG

[35] Die zitierte Literatur bezieht sich überwiegend noch auf die alte Fassung des TKG, die den Begriff der Ge-

Literatur bezüglich der Vorgängernorm (§ 6 TDG), die die Einschränkung auf „in der Regel gegen Entgelt angebotene Telemedien“ noch nicht enthielt, vertreten, dass dafür keine Gewinnerzielungsabsicht, sondern lediglich die Nachhaltigkeit des Angebots erforderlich ist ([Schm05, S. 2143], Spindler in [SpSG04, Rn. 7 zu § 6 TDG]). Die Gegenmeinung ([Woit03, S. 872], [Stic04, S. 112 f.], [Webe02, Abs. 14, 16]) erkennt diese Argumentation an, nimmt aber eine teleologische Reduktion der Norm vor: Argumentiert wurde beispielsweise, § 6 TDG sei eine verbraucherschützende Norm im Zusammenhang mit elektronischem Geschäftsverkehr, seine Anwendbarkeit auf private Homepages daher vom Gesetzgeber nicht intendiert [Woit03, S. 872]. Die Befürworter dieser Ansicht konnten auf die Gesetzesbegründung verweisen: In der Begründung zum Regierungsentwurf zur alten Fassung des TDG wird § 6 als verbraucherschützende Vorschrift bezeichnet und mit der räumlichen Trennung zum „möglichen Vertragspartner“ begründet [Bund97, S. 21]. Zwar wurde § 6 TDG später geändert, doch diente diese Änderung ausweislich der Gesetzesbegründung [Bund01b, S. 21] dazu, zusätzliche Informationsverpflichtungen zu ergänzen. Spindler [SpSG04, Rn. 7 zu § 6 TDG] indes geht davon aus, dass die Norm der Sicherung von Transparenz und Rechteverfolgung dient – ein Schutzzweck, der auch für private Angebote zutrifft. Auch diese Auffassung kann sich auf die Gesetzesbegründung stützen, nach der die Vorschrift – im Einklang mit der Definition des § 3 Nr. 10 TKG – „nur für geschäftsmäßige Angebote, die aufgrund einer nachhaltigen Tätigkeit mit oder ohne Gewinnerzielungsabsicht abgegeben werden“. Da der Gesetzesbegründung also keine eindeutige Aussage zu entnehmen war, was nachhaltige private Tätigkeiten angeht, war (trotz der gerechtfertigten Bedenken, Privatnutzern die Preisgabe ihrer Kontakdaten abzuverlangen) – dem Wortlaut des § 6 TDG und der Definition der Geschäftsmäßigkeit des § 3 Nr. 10 TKG folgend – vom Vorliegen einer Kennzeichnungspflicht auch bei privaten Anbietern auszugehen. Erst die Klarstellung im TMG, dass die Informationspflichten lediglich „in der Regel gegen Entgelt angebotene Telemedien“ betrifft, schafft hier Klarheit. Dennoch kann in manchen Fällen weiter ein Problem bestehen: In einem Peer-to-Peer-System – insbesondere in dem hier betrachteten selbstorganisierenden Empfehlungssystem – ist jeder Teilnehmer auch Diensteanbieter (vgl. Abschnitt 3.5.2, S. 48 ff.). Falls dieses Angebot nachhaltig ist und einen „in der Regel gegen Entgelt“ angebotenen Dienst betrifft, ist demnach eine Kennzeichnungspflicht nach § 5 TMG gegeben. Dienste, die in der Regel gegen Entgelt angeboten werden, sind bei Peer-to-Peer-Systemen zwar bislang die Ausnahme, sie sind aber durchaus denkbar. Doch wann ist die Nachhaltigkeit gegeben? Im Gegensatz zu einem typischen Web-Angebot ist ein Peer-to-Peer-Knoten oft nicht ständig mit dem Netz verbunden; gerade ihre Dynamik zeichnet solche Netze aus. Doch erfordert Nachhaltigkeit auch keine ständige Verfügbarkeit; wer sich nur einmalig für kurze Zeit mit dem Netz verbindet, handelt nicht nachhaltig, doch wer über Monate hinweg täglich für einige Stunden dem System beitritt, bei dem kann ein nachhaltiges Angebot

schäftsmäßigkeit in § 3 Nr. 5 TKG definierte; jedoch hat sich der Wortlaut der entsprechenden Bestimmung nur geringfügig und ohne Auswirkung auf die Definition der Geschäftsmäßigkeit geändert.

wohl ohne weiteres angenommen werden. Dies führt zu dem wenig befriedigenden Ergebnis, dass nach den Regelungen des TMG ein Teil der Netzteilnehmer – nämlich die Gruppe der nachhaltigen Anbieter, deren Dienst jedoch nur bei Betrachtung der zeitlichen Komponente vom Dienst der anderen Teilnehmer zu unterscheiden ist – zur Angabe von Name und ladungsfähiger Anschrift verpflichtet ist. Problematisch ist dies deshalb, weil diese Knoten gleichzeitig auch Nutzer des Dienstes sind und in dieser Rolle durch die Datenschutzgesetzgebung gerade vor der Preisgabe personenbezogener Daten geschützt werden sollen.

Verschärft wird diese Problematik durch die im März 2007 neu in den Rundfunkstaatsvertrag (RStV) aufgenommenen Regelungen über Telemedien. Nach § 55 Abs. 1 RStV müssen Anbieter von Telemedien, die „nicht ausschließlich persönlichen oder familiären Zwecken dienen“, ihren Namen und Anschrift ständig verfügbar halten. Die Pflicht zur Anbieterkennzeichnung wird somit über die Regelung des § 5 TMG hinaus ausgedehnt. Fraglich ist, ob bei Nutzern eines Peer-to-Peer-Empfehlungssystems von „persönlichen Zwecken“ ausgegangen werden kann. Die Begründung der Norm will auch den „gelegentlichen privaten wirtschaftlichen Geschäftsverkehr“ von der Pflicht zur Anbieterkennzeichnung ausnehmen [RStV07, S. 17]. Wiederum ist bislang davon auszugehen, dass in aller Regel ein Nutzer des Peer-to-Peer-Systems dies zu persönlichen Zwecken tut und somit auch in seiner Rolle als Diensteanbieter dieser Pflicht zur Anbieterkennzeichnung nicht unterworfen ist; Ausnahmen sind jedoch auch hier denkbar. Dies gilt insbesondere mit Blick auf mögliche zukünftige Entwicklungen.

Es bestehen grundsätzlich zwei Möglichkeiten zur Lösung des Konflikts zwischen der Pflicht zur Anbieterkennzeichnung und dem Schutz personenbezogener Daten:

- Zum einen könnten auf technischem Weg Anbieter- und Nutzerrolle getrennt werden – beispielsweise durch Verwendung unterschiedlicher Pseudonyme für beide Rollen. Es ist jedoch fraglich, ob dies für Peer-to-Peer-Systeme im allgemeinen Fall möglich ist, da Reputationssysteme eine Verknüpfung beider Funktionen benötigen oder bereits die Adressierung der Knoten zu dieser Verknüpfung beitragen kann.

- Ist keine technische Trennung möglich, so ergibt sich eine unbefriedigende Situation: Der Diensteanbieter hat dem Nutzer nach § 13 Abs. 6 TMG grundsätzlich die anonyme oder pseudonyme Nutzung des Dienstes zu ermöglichen. Andererseits führt die Verpflichtung zur Anbieterkennzeichnung dazu, dass der selbe Nutzer, übt er – wie in Peer-to-Peer-Systemen vorgesehen – auch die Funktion eines Diensteanbieters aus, sein Pseudonym aufdecken muss. § 13 Abs. 6 TMG enthält jedoch eine Einschränkung: Die Pflicht, eine zumindest pseudonyme Nutzung zu ermöglichen, gilt nur, soweit dies „soweit dies technisch möglich und zumutbar ist“. Im Zusammenhang mit der Pflicht zur Anbieterkennzeichnung ist dies aber gerade nicht der Fall. Zudem gilt zu beachten, dass gerade nicht der jeweilige Diensteanbieter die pseudonyme Nutzung unmöglich macht; vielmehr findet die Aufdeckung durch den Nutzer selbst statt.

Viel problematischer ist in diesem Zusammenhang jedoch die so erreichte Personenbeziehbarkeit von Nutzungsdaten durch den Diensteanbieter, der die Anbieterkennzeichnung derjenigen Nutzer, die eine solche bereithalten, jederzeit abrufen und mit den unter Pseudonym gespeicherten Daten zusammenführen kann. Der Personenbezug dieser Daten – verbunden mit der Tatsache, dass ein Teil der Nutzungsdaten zur Erstellung von Empfehlungen für andere Nutzer langfristig gespeichert werden kann – führt zu einem drastisch reduzierten Datenschutzniveau für einen Teil der Dienstnutzer. Die Einwilligung der Nutzer würde den Betrieb des Systems zwar dennoch ermöglichen; doch wäre ein solches System wohl für nachhaltige Nutzer nicht mehr attraktiv.

Im Lichte dieser Überlegungen erscheint eine einschränkendere Auslegung der Kennzeichnungspflicht des § 5 TMG geboten, als sie sich bei der reinen Betrachtung „klassischer“ Diensteanbieter, wie sie der Würdigung dieser Vorschrift in der Literatur und durch den Gesetzgeber zugrunde liegt, ergibt. Grundlage dieser restriktiven Auslegung ist das Recht auf informationelle Selbstbestimmung: Offensichtlich ist die Verpflichtung zur Preisgabe einer ladungsfähigen Anschrift ein schwerwiegender Eingriff in dieses Recht, der einer entsprechenden Rechtfertigung bedarf. Im Fall der klassischen Website, wie der Gesetzgeber sie im Auge hatte, ist diese zumindest tendenziell gegeben: Der Anbieter begibt sich in diesem Fall bewusst in die Öffentlichkeit; der Nutzer soll wissen, mit wem er es zu tun hat, beispielsweise um im Streitfall seine Interessen verfolgen zu können (vgl. Spindler in [SpSG04, Rn. 1 zu § 6 TDDSG]).

Im Fall eines selbstorganisierenden Empfehlungssystems – aber auch in anderen, vergleichbaren Anwendungsszenarien – stellt sich die Interessenlage jedoch anders dar: Zum einen ist die Eingriffsintensität deutlich erhöht, da der Diensteanbieter auch in seiner Rolle als Nutzer betroffen ist; zum anderen kann aber auch das dem Anbieter innewohnende Gefährdungspotential deutlich reduziert werden. So ist ein Empfehlungssystem auf einen engen Anwendungsbereich zugeschnitten, was auch das Spektrum möglicher Rechtsverletzungen verringert. Reputationsmechanismen können weiterhin dazu beitragen, die Wirksamkeit rechtsverletzender Äußerungen und Handlungen einzuschränken. Schließlich ist auch der Beitrag eines einzelnen Diensteanbieters zum Gesamtsystem vergleichsweise gering. All dies führt dazu, dass die Informationspflichten aus § 5 TMG einen unangemessenen und nicht gerechtfertigten Eingriff in das informationelle Selbstbestimmungsrecht aus Art. 1 Abs. 2, 2 Abs. 1 GG darstellen würden. Im Vergleich zur Regelung des § 6 TDG ist dieses Problem jedoch deutlich gemindert; nach bisherigem Stand ist nicht davon auszugehen, dass Empfehlungssysteme „in der Regel gegen Entgelt“ angeboten werden, so dass derzeit auch die Informationspflichten des § 5 TMG nicht greifen.

Bei einer Erweiterung des angebotenen Dienstes ist nicht auszuschließen, dass manche

Systemteilnehmer nach wie vor der Kennzeichnungspflicht unterliegen; dies dürfte jedoch eher einen Ausnahmefall darstellen, der zwar beim Systementwurf zu berücksichtigen ist, jedoch kein ernstes Hindernis für dessen Benutzbarkeit darstellt.

Daneben gilt festzuhalten, dass diese Problematik bei nicht gewerbsmäßig agierenden Diensteanbietern in der Praxis bislang keine Rolle gespielt hat und damit auch in Zukunft nicht zu rechnen ist. Für gewerbsmäßig handelnde Anbieter hingegen bestand die Kennzeichnungspflicht bereits im alten TDG; diese Unterscheidung ist auch durchaus sachgerecht.

Die rein technische Realisierung der Anbieterkennzeichnung ist unproblematisch; so gilt es lediglich, eine Schnittstelle für die Bereitstellung der entsprechenden Informationen vorzusehen.

3.6.8 Fazit

Im vorliegenden Abschnitt wurden die rechtlichen Anforderungen an den Datenschutz in verteilten, selbstorganisierenden Systemen betrachtet. Die Ergebnisse dieser Betrachtung sind in [Sorg07b] veröffentlicht; ob sich für Empfehlungssysteme weitere Besonderheiten ergeben, soll im nächsten Abschnitt diskutiert werden.

3.7 Besonderheiten von Empfehlungssystemen

Die bisherigen Ausführungen basieren zwar auf der Einordnung selbstorganisierender, verteilter Empfehlungssysteme als Telemedien; sie sind jedoch auf alle verteilten Systeme anwendbar, die sich als Telemedien einordnen lassen. Doch ergeben sich auch Besonderheiten dieser speziellen Anwendung? Solche Besonderheiten könnten sich ergeben aus

- der Art der Daten. Das deutsche Datenschutzrecht stellt an die Zulässigkeit der Erhebung, Verarbeitung und Nutzung „besonderer Arten personenbezogener Daten“ besondere Anforderungen (für nicht-öffentliche Stellen in § 28 Abs. 6–9 BDSG). Diese sind definiert als „Angaben über die rassische und ethnische Herkunft, politische Meinungen, religiöse oder philosophische Überzeugungen, Gewerkschaftszugehörigkeit, Gesundheit oder Sexualleben“ (§ 3 Abs. 9 BDSG). In Empfehlungssystemen werden solche Angaben allerdings in aller Regel nicht benötigt. In Einzelfällen kann es jedoch vorkommen, dass sich aus den im Empfehlungssystem verwendeten Daten mit einer gewissen Wahrscheinlichkeit eine Aussage beispielsweise über die Gesundheit eines Nutzers ableiten lässt (so beispielsweise aus empfohlener Literatur über die religiöse Überzeugung).
- dem Verwendungszweck der Daten. § 28 Abs. 3 Nr. 3 erlaubt die Übermittlung oder Nutzung personenbezogener Daten zum Zweck der Werbung sowie der Markt- und

Meinungsforschung, sofern „kein Grund zu der Annahme besteht, dass der Betroffene ein schutzwürdiges Interesse an dem Ausschluss der Übermittlung oder Nutzung hat“. Für Empfehlungssysteme ist diese Ausnahmeregelung jedoch nicht relevant. Zwar können diese durchaus auch zum Zweck der Werbung eingesetzt werden, doch beschränkt sich die gesetzliche Erlaubnis auf „listenmäßig oder sonst zusammengefasste Daten“ und gibt zudem die Art der Daten, die übermittelt werden können, vor. Konkret handelt es sich um die Angabe über die Zugehörigkeit des Betroffenen zu einer Personengruppe, eine Berufs-, Branchen- oder Geschäftsbezeichnung sowie Name, Titel, akademische Grade, Anschrift und Geburtsjahr. Diese Daten reichen jedoch in aller Regel nicht aus, um eine Empfehlung zu berechnen. Theoretisch wäre denkbar, die benötigten Informationen durch die Zuordnung zu Personengruppen zu codieren – beispielsweise die Gruppe der „Schatz im Silbersee“-Leser. § 28 Abs. 3 Nr. 3 BDSG spricht jedoch lediglich von *einer* Personengruppe. Die Übermittlung der Tatsache, dass ein Nutzer zu zahlreichen Gruppen gehört, die sich jeweils dadurch auszeichnen, jeweils ein bestimmtes Objekt gekauft zu haben, wäre durch diese Erlaubnisnorm nicht mehr gedeckt. Dies ergibt sich aus einer teleologischen Auslegung der Norm, die durch die so mögliche Codierung beliebiger Informationen über eine Person zur fast vollständigen Aufhebung der Schutzfunktion des BDSG führen würde.

- dem Umfang der erhobenen Daten. Allerdings können in einem Empfehlungssystem zwar personenbezogene Daten in großem Umfang erhoben werden, doch gilt dies zum einen auch für zahlreiche andere Anwendungen, zum anderen kennt das deutsche Datenschutzrecht keine untere Grenze für Umfang oder Bedeutung personenbezogener Daten. Dies ergibt sich bereits aus dem Volkszählungsurteil des BVerfG [Bund83].

In der Summe kann somit festgehalten werden, dass sich für selbstorganisierende, verteilte Empfehlungsdienste zwar ein hohes tatsächliches Gefährdungspotential in Bezug auf den Datenschutz ergibt, aus rechtlicher Sicht allerdings keine Unterschiede zu sonstigen selbstorganisierend und verteilt funktionierenden Telemedien besteht.

3.8 Fazit

In diesem Kapitel wurden Anforderungen von Internetnutzern an den Datenschutz analysiert und daraus ein Modell aus fünf Ebenen – Inhalts-, Empfänger-, Identitäts-, Zweck- und Kontrollebene – entwickelt. Ansätze zur Erfüllung dieser Anforderungen wurden sowohl aus technischer als auch aus juristischer Sicht diskutiert. Zusammenfassend stellt Tabelle 3.1 dar, wie sich diese Ansätze in das entwickelte Anforderungsmodell aus Abschnitt 3.1.1.1

einordnen lassen. Hierbei wird deutlich, dass das TMG insbesondere in der Kontrollebene zahlreiche Anforderungen an die technische Umsetzung von Telemedien stellt. Dennoch bestehen auf der Kontrollebene Defizite: Trotz aller gesetzlichen Regelungen, die Transparenz herstellen und dem Nutzer somit die Kontrolle über preisgegebene personenbezogene Daten geben sollen, kann er direkten Einfluss lediglich bei der Erhebung seiner Daten ausüben. Dieser Einfluss erstreckt sich auf die Inhalts- und (mit Einschränkungen) die Identitäts- sowie die Empfängerebene. Der Nutzer kann übersehen, welche Daten er preisgibt und in der Regel auch, unter welcher Identität. Welcher Empfänger die Daten durch ihn erhält, ist ebenfalls für den Nutzer kontrollierbar – nicht jedoch, an wen die Daten weitergegeben werden. Auch bezüglich des Zwecks, zu dem die Daten verarbeitet werden, muss er sich auf die Aussagen des Verwenders verlassen. Zwar enthält das BDSG in den §§ 43,44 Bußgeld- und Strafvorschriften, und eine Reihe von Vorschriften (§§ 4d–4g BDSG sowie Regelungen über die Datenschutzkontrolle in den Datenschutzgesetzen der Länder) dient ebenfalls der Kontrolle der Verarbeitung personenbezogener Daten. Doch gestaltet sich die Überprüfung von Datenverarbeitungsvorgängen in der Praxis schwierig, sodass trotz guter Ansätze im Gesetz die Kontrollebene als problematisch anzusehen ist. Auch durch weitere gesetzgeberische Maßnahmen ist hier Abhilfe kaum möglich, da Rechtsverstöße schwer zu entdecken sind.

Auch bei rechtmäßigem Handeln werden durch das Datenschutzrecht nicht alle Ebenen hinreichend abgedeckt: Wenn beispielsweise das Datenschutzrecht personenbezogene Daten auf der Kontroll- und der Zweckebene schützt, kann die Technik darüber hinausgehen, indem einerseits die Identitätsebene – durch eine geeignete Verwaltung von Pseudonymisierung und Anonymisierung – einbezogen wird, andererseits auch Pseudonyme selbst als wertvoll betrachtet und die mit ihnen verknüpften Daten begrenzt werden können.

Ein Schwerpunkt soll in dieser Arbeit auf dem Datenschutz, aber auch der Funktionalität eines Empfehlungssystems auf der Basis von Vertrauensbeziehungen liegen. Deshalb wird im Folgenden das Thema Vertrauen ausführlich erörtert.

	Anforderung	Erfüllung durch Datenschutzrecht	Anforderung des Rechts an die Technik	Technische Lösungsansätze
Inhaltsebene	Bereitschaft zur Preisgabe personenbezogener Daten abhängig von Art dieser Daten.	Im allgemeinen Datenschutzrecht (BDSG) durch Unterscheidung „besonderer Arten personenbezogener Daten“ (§ 3 Abs. 9 BDSG) und sonstiger personenbezogener Daten widergespiegelt – hierbei handelt es sich jedoch um eine nur grobe Unterteilung. Im TMG findet sich eine entsprechende Unterscheidung nicht.	—	—
Empfängerebene	Bereitschaft zur Preisgabe personenbezogener Daten abhängig von den Empfängern dieser Daten.	Das BDSG unterscheidet zwischen öffentlichen und nicht-öffentlichen Stellen als Empfänger von Daten. Einschränkungen der Übermittlung personenbezogener Daten stellen sicher, dass – mit wenigen Ausnahmen, wie sie z.B. in § 15 Abs. 5 TMG definiert sind – nur der durch den Nutzer gewünschte Empfänger diese Daten erhält.	—	Technische Maßnahmen können den Nutzer dabei unterstützen, einen vertrauenswürdigen Empfängerkreis auszuwählen; dies wird in Kapitel 4 diskutiert.

Identitäts-ebene	Bereitschaft zur Preisgabe persönlicher Daten abhängig von der Identität, mit der diese Daten verknüpft sind.	Anwendbarkeit des Datenschutzrechts nur bei personenbezogenen Daten (§ 1 Abs. 2 BDSG; § 12 Abs. 1 TMG)	Pflicht des Diensteanbieters • die anonyme oder pseudonyme Nutzung des Dienstes zu ermöglichen (§ 13 Abs. 6 TMG) • Nutzungsprofile lediglich unter Verwendung von Pseudonymen zu erstellen, die nicht mit Daten über den Pseudonymträger zusammengeführt werden dürfen (§ 13 Abs. 4 Nr. 6 TMG)	• Identitätsmanagement (Verwaltung der verschiedenen Identitäten einer Person bei der Interaktion mit verschiedenen Transaktionspartnern oder Diensteanbietern) • Technische Ansätze zur Anonymisierung in Netzen (vgl. Abschnitte 3.3.1,3.3.2) Erfüllung nur der juristischen Anforderungen reicht u.U. nicht aus: Auch Pseudonyme ohne Personenbezug können für den Nutzer wertvoll und schützenswert sein.
Zweck-ebene	Bereitschaft zur Preisgabe persönlicher Daten hängt von dem Zweck ab, zu dem diese Daten verwendet werden sollen.	Zweckbindungsgrundsatz der §§ 28 Abs. 2 BDSG, 12 Abs. 2 TMG	—	—

Kontroll-ebene	Kontrolle über preisgegebene Daten; beinhaltet • Transparenz der Preisgabe von Daten • Einflussmöglichkeit auf bereits preisgegebene Daten	Der überwiegende Teil der Regelungen zum Systemdatenschutz in § 13 Abs. 4–7 TMG sowie die Verpflichtung zur Anbieterkennzeichnung des § 5 TMG (vgl. Abschnitt 3.6.7 dienen der Herstellung von Transparenz; die Regelung des § 13 Abs. 7 TMG (Pflicht zur Auskunftserteilung) soll darüber hinaus den Einfluss auf bereits preisgegebene Daten ermöglichen.	Die Regelungen zum Systemdatenschutz bedürfen einer technischen Umsetzung, die dem Nutzer insbesondere Transparenz darüber gibt, welche Daten übermittelt werden.	Auch über die gesetzlichen Anforderungen hinaus bestehen zahlreiche Ansätze, die die Transparenz bei der Verwendung personenbezogener Daten erhöhen, beispielsweise Filter-Tools und Policy-Tools (vgl. Abschnitt 3.3).

Tabelle 3.1: Erfüllung von Datenschutzanforderungen

Kapitel 4

Vertrauen

Das Phänomen Vertrauen spielt in Bezug auf das zu entwickelnde Empfehlungssystem in vielerlei Hinsicht eine große Rolle. Dies betrifft beispielsweise das Vertrauen in Empfehlungen anderer Marktteilnehmer, das sich wiederum auf das Vertrauen in diese Personen stützt. Notwendig für die Einführung eines technischen Systems ist aber auch immer das Vertrauen der Nutzer in das System.

Bevor auf die konkrete Anwendung von Vertrauen und Vertrauensbeziehungen in Peer-to-Peer-Systemen – und, in der Folge, in Peer-to-Peer-basierten Empfehlungssystemen – eingegangen wird, wird das Thema zunächst allgemeiner betrachtet. Zunächst wird dazu die sozialwissenschaftliche Sicht herangezogen. Eng damit verbunden ist eine ökonomische Betrachtung des Themas. Das Verhältnis zwischen Vertrauen und Recht wird im Anschluss erörtert. Kernpunkt des Kapitels ist die Vertrauenserzeugung in Netzen, für die Reputationssysteme als wesentliches Konzept identifiziert werden. Herausforderungen für Reputationssysteme werden aus Sicht der Informatik und des Rechts betrachtet. Aus den Ergebnissen, die für Reputationssysteme erzielt werden, lassen sich Folgerungen für Empfehlungssysteme ableiten.

Den Abschluss des Kapitels bildet ein Bogenschlag vom Konzept Vertrauen zur Vertraulichkeit. Ein Verfahren, das auf Grundlage eines Vertrauensmodells auch eine Entscheidung über die Preisgabe vertraulicher Informationen treffen kann, wird vorgestellt.

4.1 Grundlagen

Den ersten Schritt der Diskussion des Themas Vertrauen bildet notwendigerweise die Suche nach einer Definition des Begriffs. Eine einheitliche Definition des Begriffs existiert nicht[1]; Einigkeit besteht lediglich darüber, dass Vertrauen mit einem Risiko behaftet ist (vgl. [Will93, S. 463], [Cole90, S. 91], [Levi01, S. 15922]). Als Ausgangspunkt der Betrachtung von Vertrauen in dieser Arbeit können die Definitionen von Esser [Esse00], Sztompka [Szto01] und Gambetta [Gamb88] dienen:

[1] Definitionen finden sich beispielsweise in [Brai98, S. 47], [Esse00, S. 73], [Szto01, S. 15913], [Cole90, S. 91].

- Esser [Esse00, S. 73] definiert Vertrauen als „eine generalisierte Erwartung, dass eine Vorleistung nicht ausgenutzt wird“. Dabei ist der Erwartende sich aber bewusst, dass ein Risiko besteht: Die Vorleistung *kann* eben doch ausgenutzt werden.
- Auch Sztompka [Szto01, S. 15913] betont diesen Aspekt der Ungewissheit, indem er Vertrauen als Wette über zukünftige, ungewisse Handlungen Anderer bezeichnet.[2] Wer vertraue, der handle, als kenne er die Art und Weise, in der andere sich verhielten. Implizit enthält auch diese Definition das Konzept einer Vorleistung, denn bei einer Wette muss ein Einsatz erbracht werden.
- Gambetta [Gamb88, S. 217] definiert Vertrauen als „[...] subjektive Wahrscheinlichkeit, mit der ein Agent[3] davon ausgeht, dass ein anderer Agent oder eine Gruppe von Agenten eine bestimmte Handlung ausführt, bevor er die Handlung beobachten kann (oder unabhängig davon, ob er diese jemals wird beobachten können), und zwar in einem Kontext, in dem seine eigenen Handlungen davon beeinflusst werden.“[4]

Einer Verfeinerung bedürfen diese Definitionen bezüglich der zeitlichen Komponente: Das Verhalten, auf das vertraut wird, muss nicht *nach* der Vertrauenshandlung geschehen. Wichtig ist allein, dass der Vertrauende das Verhalten, auf dass er sich verlässt, nicht beobachten kann.

In einem Versuch, die Definitionen zusammenzufassen, soll in dieser Arbeit unter dem Begriff Vertrauen verstanden werden:

Definition 4. *Vertrauen* ist die Annahme, dass das bewusste, risikobehaftete Eingehen einer Abhängigkeit vom (zum Entscheidungszeitpunkt nicht beobachtbaren)

- Verhalten einer Person, Sache, Institution oder eines Systems
- oder Vorliegen eines bestimmten Sachverhalts

den eigenen Interessen dient. Eine Person, Sache, Institution oder ein System, dem vertraut wird, heißt *Vertrauensnehmer*. Der Vertrauensnehmer ist *vertrauenswürdig*, wenn es sich dieser Annahme entsprechend verhält.

Wichtig ist an dieser Definition, dass Vertrauen sich nicht nur auf Personen beziehen kann, sondern beispielsweise auch auf den Staat oder andere Institutionen.

Zur Abgrenzung zwischen Vertrauen und Sicherheit bleibt anzumerken, dass beide Konzepte sich ergänzen. Ist ein System aus Sicht seiner Nutzer sicher, dann besteht kein Risiko; Vertrauen ist also nicht notwendig. Zum einen muss aber selbst ein objektiv sicheres System nicht auch als sicher wahrgenommen werden – in vielen Fällen sind die Nutzer schlicht nicht in der Lage, die Sicherheit des Systems objektiv einzuschätzen, weshalb Vertrauen notwendig ist. Zum anderen kann es vorkommen, dass ein System nicht oder nur unter prohi-

[2] „Trust is a bet about the future contingent actions of others.“

[3] Der Begriff wird in diesem Kontext als Synonym für „Akteur“ verwendet.

[4] „a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor such action (or independently of his capacity ever to be able to monitor it) and in a context in which it affects his own action.“

bitiv hohem Aufwand abgesichert werden kann. Auch in diesem Fall erlaubt das Konzept des Vertrauens den wirtschaftlich sinnvollen Einsatz des Systems.

Wer beispielsweise vertrauliche Daten per E-Mail an einen bestimmten Empfänger übermittelt, der kann (auf den ersten Blick) zwar durch Verschlüsselung sicherstellen, dass sie nur von diesem Empfänger gelesen werden können. Vertrauen braucht er dennoch, und zwar bezüglich der Art, wie der Empfänger mit den Daten umgeht. Bei genauerer Betrachtung stellt man fest, dass Vertrauen in diesem Szenario sogar in vielerlei Hinsicht notwendig ist:

- Vertrauen in das eigene Computersystem, das die Verschlüsselung vornimmt, einschließlich der verwendeten Software.
- Vertrauen in die Authentizität des verwendeten Schlüsselmaterials.
- Vertrauen, dass die verwendeten kryptographischen Algorithmen nicht in absehbarer Zeit gebrochen werden können.
- Vertrauen nicht nur in den guten Willen, sondern auch in die Kompetenz des Empfängers, die Daten sicher zu halten.

4.1.1 Vertrauen aus soziologischer Sicht

Grundlegende Eigenschaften des Vertrauens als gesellschaftliches Phänomen werden in der Soziologie untersucht, die deshalb Ausgangspunkt der vorliegenden Betrachtung sein soll. Eine Einführung in die soziologische Sicht auf Vertrauen wird in [Levi01] gegeben. Ausgehend von dieser Arbeit werden im Folgenden mögliche Motivationen für Vertrauen diskutiert und die Messbarkeit von Vertrauen untersucht.

4.1.1.1 Warum Vertrauen?

[Levi01, S. 15923] identifiziert vier mögliche Motivationen, zu vertrauen:

- *Veranlagung.* In der Literatur ist anerkannt, dass Menschen unterschiedlich stark geneigt sind, zu vertrauen. Ob und inwieweit dies am Charakter, Prägung in der Kindheit oder dem sozialen Umfeld liegt, ist jedoch umstritten.
- *Moral.* Einerseits steht *Vertrauenswürdigkeit* in engem Zusammenhang mit moralischem Verhalten; andererseits wird auch das Vertrauen selbst von der Gesellschaft grundsätzlich (wenn auch nicht undifferenziert) als positiv betrachtet. Dies ist auch einer der Gründe, warum Vertrauen von der Rechtsordnung geschützt wird (daneben treten ökonomische Gründe – siehe dazu Abschnitt 4.2.1).
- *Soziale Gründe.* Eng damit verknüpft sind soziale Gründe: Das Leben in der Gemeinschaft erfordert Vertrauen. Wer nicht vertraut oder sich nicht als vertrauenswürdig erweist, dessen sozialer Status ist gefährdet. Hinweis auf diese Motivation des Vertrauens ist die unterschiedliche Vertrauensneigung in verschiedenen Gesellschaften (siehe nur [YaYa94, S. 130]).

- *Rationale Erwägungen.* Im Zentrum der wirtschaftswissenschaftlichen Betrachtung stehen rationale Gründe, die zu Vertrauen führen. Ist in einer Entscheidungssituation der erwartete Nutzen aus der Entscheidung, zu vertrauen, größer als derjenige aus der Entscheidung, nicht zu vertrauen, so wird der rational denkende Mensch sich für das Vertrauen entscheiden.[5]

4.1.1.2 Messbarkeit von Vertrauen

Die oben aufgestellte Definition ermöglicht lediglich die Unterscheidung, ob man vertraut oder nicht. Will man Abstufungen von Vertrauen messen, so kann auf die Definition von Gambetta [Gamb88, S. 217] zurückgegriffen werden. Unterschieden werden kann dann die subjektive (wahrgenommene) Wahrscheinlichkeit, mit der die Annahme, der Interaktionspartner sei vertrauenswürdig, zutrifft.

Diese explizit anzugeben, ist jedoch meist nicht möglich. Die Aussagekraft entsprechender Umfragen ist deshalb begrenzt; so wird nicht notwendigerweise ausschließlich Vertrauen gemessen, und verschiedene Umfragen können i.A. nicht miteinander verglichen werden. [Levi01, S. 15924 f.]

In der Literatur finden sich verschiedene Ansätze, das Problem zu umgehen; so wird der Geldbetrag, den man einer Person anzuvertrauen bereit ist, als Maß verwendet [GLSS00] – ein Ansatz, der sich jedoch nicht beliebig verallgemeinern lässt, da nicht in allen Domänen ein monetäres Äquivalent gefunden werden kann. In vielen Fällen wird daher auf die Zuweisung einer expliziten Bedeutung zu verschiedenen Stufen einer Vertrauensskala verzichtet, so beispielsweise im PGP-Web-of-Trust [Abdu97], das lediglich unscharfe Bezeichnungen wie „marginal trust“ kennt. Marsh [Mars94, Abschnitt 2.3] schlägt die Verwendung von reellen Zahlen im Intervall [-1;1] vor, die für menschliche Nutzer durch verständliche Bezeichner (wie „blindes Vertrauen“) ergänzt werden. Im Rahmen dieser Arbeit soll das Konzept, Vertrauen als Zahlenwert abzubilden, zur Vereinfachung übernommen werden; dabei wird das Intervall [0;1] gewählt. Eine Notwendigkeit zur Modellierung von Misstrauen (durch negative Zahlenwerte) besteht im Rahmen dieser Arbeit nicht.

4.1.1.3 Domänenabhängigkeit von Vertrauen

In zahlreichen Anwendungen wird sich die Domänenabhängigkeit von Vertrauen als wichtiger Aspekt herausstellen. Vertrauen lässt sich nicht mit der Aussage „A vertraut B“ darstellen, sondern nur als „A vertraut darauf, dass B die Handlung Z vornimmt“ [Jone01, S. 15918] – oder „A vertraut B mit Bezug auf Domäne D“.

[5] Ist der Entscheider nicht risikoneutral, so sind Risiko-Nutzenfunktionen zugrunde zu legen.

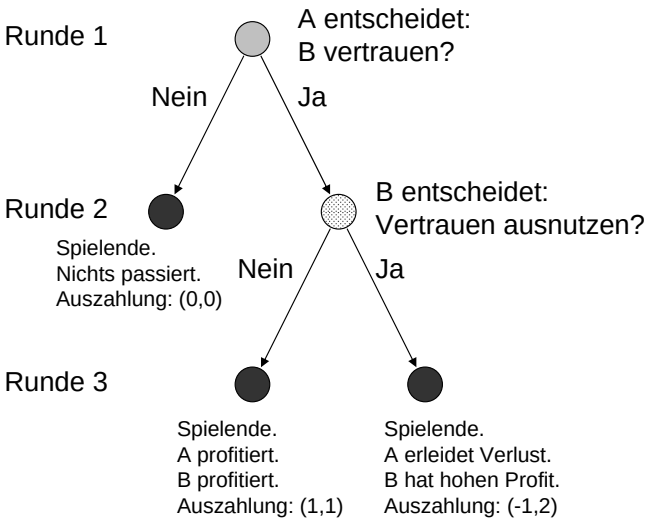

Abbildung 4.1: Das Vertrauensspiel (vgl. [Jone01], [Dasg88, S. 59 ff.])

4.1.2 Vertrauen aus wirtschaftswissenschaftlicher Sicht

Aus wirtschaftswissenschaftlicher Sicht existieren zahlreiche Beiträge über das Phänomen Vertrauen. Ausgangspunkt einer Untersuchung kann die Spieltheorie sein, denn die Betrachtung einer Vertrauensentscheidung als Spiel ermöglicht die Einordnung zahlreicher mit dieser Entscheidung verbundener Einflussfaktoren.

4.1.2.1 Spieltheorie

Der oben (S. 86) erarbeiteten Definition von Vertrauen folgend, muss ein Spiel, das eine Vertrauensentscheidung abbilden soll, im Wesentlichen zwei Aspekte modellieren: Zum Einen muss eine risikobehaftete Entscheidung getroffen werden, d.h. das Verhalten des jeweils anderen Spielers steht nicht sicher fest. Zum Anderen geht ein Spieler, der vertraut, eine Abhängigkeit ein. Sobald er sich entschieden hat, zu vertrauen, kann er den Ausgang des Spiels nicht mehr bestimmen. Diese Entscheidung liegt vielmehr in der Hand des anderen Spielers.

Das Spiel in Abbildung 4.1 genügt diesen Charakteristika: In der ersten Runde entscheidet A, ob er B vertrauen soll oder nicht. Tut er es nicht, profitiert keiner der Akteure aus der Interaktion, d.h. beide erhalten eine Auszahlung von null. Entscheidet A jedoch, B zu vertrauen, so kann B in der zweiten Runde entscheiden, ob er dieses Vertrauen ausnutzt oder sich als vertrauenswürdig erweist. Im ersten Fall verliert A durch diese Entscheidung (Auszahlung -1), wohingegen B profitiert (Auszahlung 2). Im zweiten Fall profitieren beide Spieler (Auszahlung 1 für A und B); der Gewinn des Spielers B ist jedoch geringer als im ersten Fall. Dies ist auch der Grund, warum ein rationaler Spieler B sich in der zweiten Runde dafür entscheiden würde, das Vertrauen seines Interaktionspartners A auszunutzen.

Da A in Kenntnis dieser Tatsache B in der ersten Runde kein Vertrauen entgegenbringt, kommt es jedoch nicht zu dieser Entscheidung. Statt des (für beide Seiten vorteilhaften) Ausgangs, in dem beide Spieler die Auszahlung 1 erhalten, wird also überhaupt kein Gewinn realisiert (beispielsweise kommt keine Geschäftsbeziehung zustande). Diese einfache Betrachtung zeigt auch, dass durch Vertrauen Effizienzgewinne realisiert, also in manchen Fällen alle Beteiligten besser gestellt werden können.

Zu beachten gilt allerdings, dass die in der Analyse des Spiels unterstellte vollständige Information der Teilnehmer in der Realität nicht gegeben ist. Bestenfalls kennt ein Spieler seine eigenen Auszahlungen in Abhängigkeit vom Ausgang des Spiels. Bezüglich möglicher Auszahlungen des anderen Spielers ist er jedoch auf seine persönliche Einschätzung angewiesen. Das heißt beispielsweise, dass Spieler A in der ersten Runde nicht weiß, wie die Auszahlungen von Spieler B in der dritten Runde sich gestalten. Daher weiß Spieler A ebenfalls nicht, wie Spieler B sich in der zweiten Runde entscheiden wird. Spieler A wird jedoch zu einer subjektiven Einschätzung der Wahrscheinlichkeit kommen können, dass Spieler B das in ihn gesetzte Vertrauen erfüllt. Die Beeinflussung dieser Einschätzung ist ein Kernproblem im Umgang mit Vertrauen, das sich auch bei Vertrauensfragestellungen in Rechnernetzen herausstellen wird.

4.1.2.2 Sonstige wirtschaftswissenschaftliche Forschung

Neben der Spieltheorie befassen sich auch andere Forschungsgebiete der Wirtschaftswissenschaften mit dem Phänomen Vertrauen. Hierzu zählen insbesondere

- die Organisationswissenschaft. Forschungsgegenstand sind hier die Auswirkungen von Vertrauen auf die Organisation innerhalb von Unternehmen sowie im Verhältnis verschiedener Unternehmen (siehe z.B. [ZaMP98]).
- die Neuroökonomik [KePl05], die an der Schnittstelle zwischen Gehirnforschung und Wirtschaftswissenschaften den Ursprung menschlichen Verhaltens untersucht. Allerdings finden sich in der Literatur zur Neuroökonomik noch keine belastbaren Ergebnisse zum Thema Vertrauen.

Neben Soziologie und Wirtschaftswissenschaften, die sich mit den Grundlagen des Vertrauens beschäftigen, spielt das Phänomen Vertrauen auch in der Rechtswissenschaft eine große Rolle.

4.2 Vertrauen im deutschen Recht

Bevor der Umgang mit Vertrauen im deutschen Recht untersucht wird, soll jedoch zunächst das Verhältnis *zwischen* Vertrauen und Recht betrachtet werden. Darmstaedter [Darm48, Sp. 433] vertritt die Auffassung, das Recht wolle Vertrauen überflüssig machen: Indem es Normen und Sanktionen beim Verstoß gegen diese Normen aufstelle, wolle es erreichen,

	Situation trust	**Behaviour trust**
Strafrecht	Verbotsirrtum (§ 17 StGB)[7]	Schutz vertraulicher Information (z.B. §§ 201, 203 StGB)
Zivilrecht: Direkter Schutz	Öffentliche Register (z.B. Handelsregister)	Vertrauenshaftung
Deliktsrecht	—	Schadensersatzansprüche (§§ 823, 826 BGB)

Tabelle 4.1: Beispiele für den Schutz von Vertrauen in verschiedenen Rechtsgebieten

dass der Vertrauende selbst dann nicht benachteiligt wird, wenn sein Vertrauen sich als nicht gerechtfertigt herausstellt. Die Rechtsordnung könne damit allerdings nur den fehlenden guten Willen zur Leistung, nicht aber das fehlende Können ersetzen. Der Vertrauensbegriff wird durch diese Argumentation beschränkt auf das Vertrauen, dass eine Person eine Leistung erbringen wird – dies hängt vermutlich damit zusammen, dass es sich hierbei um eine der Rechtsdurchsetzung besonders zugängliche Art des Vertrauens handelt.

Darmstaedters Auffassung ist nur eingeschränkt zuzustimmen. In der Tat kann die Rechtsordnung eher auf den Leistungswillen als auf das Leistungsvermögen einwirken. Sie kann jedoch in der Regel das Vertrauen nicht ersetzen. So besteht stets das Risiko, dass der Leistungspflichtige sich der Rechtsordnung entzieht oder das ihm entgegengebrachte Vertrauen in einer nicht sanktionierten Art und Weise missbraucht.

Die Rechtsordnung kann also Vertrauen begünstigen, indem das damit verbundene Risiko reduziert wird. Statt eines Substituts ist sie folglich oft eine Voraussetzung für Vertrauen. Sie hilft somit in vielen Fällen, die ökonomischen Vorteile, die sich durch Vertrauen ergeben können, auch tatsächlich zu realisieren.

Wie sich diese Unterstützung von Vertrauen durch das Recht konkret darstellt, wird von Memmo et al. [MeSC03] untersucht; dieser Beitrag diskutiert das Verhältnis zwischen Vertrauen und Recht am Beispiel des italienischen Rechts – die Ergebnisse lassen sich aber ohne weiteres auch auf die Rechtslage in Deutschland übertragen. Memmo et al. unterscheiden zwischen verschiedenen Arten von Vertrauen.[6]

- Vertrauen, dass eine bestimmte (rechtlich relevante) Situation vorliegt (*Situation trust*). Das Vertrauen in das Vorliegen einer Situation ist indes nicht Gegenstand der vorliegenden Arbeit und wird daher nicht weiter betrachtet.
- Vertrauen in das Verhalten eines Anderen (*Behaviour trust*).

[6] Weiterhin findet sich eine Unterteilung in Vertrauen im engeren Sinne (*core trust*) und Sich-Verlassen (*reliance*), auf die an dieser Stelle jedoch nicht eingegangen werden soll. Stattdessen wird in dieser Dissertation nur Vertrauen im engeren Sinne betrachtet.

Tabelle 4.1 zeigt einige Beispiele für beide Formen des Schutzes von Vertrauen in verschiedenen Rechtsgebieten auf. Im Folgenden soll jedoch nur noch das „Behaviour Trust“ betrachtet werden. Relevant wird das Verhältnis des Behaviour Trust zum Recht in drei Bereichen:

- Vertrauen aufgrund gesetzlicher Regelungen, die das Verhalten anderer beeinflussen. Hierzu gehören beispielsweise
  - das Strafrecht. Als Beispiel kann das Vertrauen eines Patienten herangezogen werden, dass sein Arzt keine vertraulichen Informationen über seine Gesundheit preisgibt. Einer der Gründe für dieses Vertrauen kann darin liegen, dass dem Arzt eine Strafe droht, falls er solche vertraulichen Informationen offenbart (§ 203 I Nr. 1 StGB).
  - das Zivilrecht, insbesondere das Deliktsrecht. Als Beispiel kann § 823 Abs. 2 BGB gelten. Besteht ein Gesetz zum Schutz einer Person, so kann diese Person von der Einhaltung des Gesetzes ausgehen. Wird das Gesetz nicht eingehalten, so kann der Verletzte Schadensersatz verlangen. Diese Regelung liefert einerseits einen Anreiz zur Einhaltung des Schutzgesetzes; andererseits wird selbst bei einer eventuell doch auftretenden Verletzung ein Ausgleich des entstandenen Schadens geschaffen. In Abbildung 4.1 bedeutet dies, dass sowohl die Wahrscheinlichkeit nicht vertrauenswürdigen Verhaltens von B als auch der in diesem Fall entstehende Schaden für A sinkt.
- Direkter Schutz von Vertrauen durch das Recht. Dieser Schutz wird in Abschnitt 4.2.1 diskutiert.
- Gesetzliche Regelungen, die sich mit vertrauensbildenden Verfahren außerhalb des Rechts beschäftigen. Reputationssysteme als wichtigstes Beispiel werden in Abschnitt Abschnitt 4.3.5 (S. 101) diskutiert.

4.2.1 Direkter Schutz von Vertrauen durch das Recht

Abgesehen von Normen, die nur indirekt vertrauensfördernd wirken, gibt es auch Regelungen, die einen direkten Schutz des Vertrauens einer Person in das Verhalten einer anderen begründen. Vertraut eine Person A darauf, dass eine Person B sich in einer bestimmten Weise verhält, und weicht Person B von diesem erwarteten Verhalten ab, führt dies zu einem (zumindest teilweisen) Ersatz des Schadens, den A dadurch erleidet. Dieser Schutz liegt im Vertrauen von A begründet, nicht in einer allgemeinen Ablehung des Verhaltens von B durch die Rechtsordnung. Wird der Ersatz des Schadens durch B geleistet, so spricht man von *Vertrauenshaftung*.

[7] Die Norm schützt das Vertrauen in die Rechtmäßigkeit einer Tat, auch, wenn diese Tat strafbar ist. Dieser Schutz ist aber auf unvermeidliche Irrtümer über die Rechtmäßigkeit beschränkt.

Konkret findet sich die Vertrauenshaftung beispielsweise in den folgenden Fällen (vgl. dazu [ScOt05, S. 525–528]):

- Das Stellvertretungsrecht. §§ 170, 171 Abs. 2, 172 Abs. 2, 173 und 179 Abs. 1 BGB schützen das Vertrauen einer Person in die Vertretungsmacht einer anderen. Die §§ 170, 171 Abs. 2 und 172 Abs. 2 BGB regeln das Fortbestehen der Vertretungsmacht, die einem Dritten auf eine bestimmte Art und Weise bekanntgegeben wurde; eine Ausnahme dieses Fortbestands wird in § 173 BGB geregelt. Vertraut ein Dritter auf einen Vertreter, der keine Vertretungsmacht hat, so wird er durch § 179 Abs. 1 BGB geschützt.
- Die Erklärungshaftung. Ficht eine Person eine von ihr abgegebene Erklärung an, so begründet dies einen Anspruch desjenigen, der auf die Gültigkeit der Erklärung vertraut, auf Ersatz des Vertrauensschadens (§ 122 BGB). Auch § 179 Abs. 1 BGB begründet (im bereits erwähnten Stellvertretungsrecht) die Haftung für eine Erklärung.
- Die Lehre vom Rechtsschein. Ein Dritter kann sich auf das Bestehen einer bestimmten Rechtslage (den Rechtsschein), auf das er vertraut hat und vertrauen durfte, dann berufen, wenn eine Person in einer zurechenbaren Weise den Anschein erweckt hat, dass diese Rechtslage bestehe [Cana71, S. 526–528].
- Widersprüchliches Verhalten. Dieses Konzept gründet sich auf § 242 BGB. Vertraut jemand in einer schutzwürdigen Weise auf das Verhalten einer anderen Person, wird der Vertrauende geschützt, wenn die andere Person sich zu seinem früheren Verhalten in Widerspruch setzt. Der Vertrauende kann sich dann auf das frühere Verhalten berufen [ScOt05, S. 528]. Allerdings kann dieser Schutz nicht so weit gehen, dass eine Person ihr Verhalten überhaupt nicht mehr ändern kann.

Doch warum schützt das Recht in diesen Fällen das Vertrauen? Aus Sicht der Ökonomie haben Schäfer und Ott [ScOt05, S. 517] vier Voraussetzungen für einen Schutz des Vertrauens durch das Recht identifiziert:

- *Asymmetrie von Informationskosten.* Kann eine der beteiligten Parteien unter geringerem Kostenaufwand eine Information beschaffen als die andere Partei, so kann es sinnvoll sein, die Partei zu schützen, die auf diese Information vertraut. Dies gilt insbesondere, wenn die Information das zukünftige Verhalten der erstgenannten Partei betrifft.
- *Gesamtgesellschaftliche Produktivität der Information.* Die Beschaffung einer Information ist dann sinnvoll, wenn – gesamtgesellschaftlich gesehen – die Kosten der Informationsbeschaffung unter dem aus der Information entstehenden Nutzen liegen. Deshalb sollte das Vertrauen der anderen Partei in die Partei, die die Information erhält, geschützt werden. Beispielsweise lohnt sich die Beschaffung der Information, ob eine Person Vertretungsmacht für einen Dritten hat: Der Schaden, den eine Partei durch

Vertrauen auf die Gültigkeit eines Vertrages erleiden kann, wird in vielen Fällen höher sein als der Nutzen der Person, die diese Information nicht bereitstellt, obwohl sie dies könnte. Aus diesem Grund wird das Vertrauen des Vertragspartners geschützt.

- *Die Existenz einer Vertrauensprämie.* Die Person, der vertraut wird, und die somit ggf. auch haftbar gemacht wird, muss für dieses erhöhte Risiko und die dadurch entstehenden Kosten auch einen Ausgleich erhalten – die sogenannte Vertrauensprämie. In einem Markt wird diese Prämie durch den Marktmechanismus erreicht; wird eine Information hingegen lediglich als Gefälligkeit zur Verfügung gestellt, gibt es auch keine Vertrauensprämie, und ein Schutz des Vertrauens wäre nicht angebracht. Eine ausführliche Diskussion des Konzepts der Vertrauensprämie findet sich in [ScOt05, S. 519–522].

- *Das Verhältnis von Vertrauens- und Opportunismusprämie.* Ein rechtlicher Schutz des Vertrauens ist nur dann nötig, wenn die Partei, der vertraut wird, einen höheren Nutzen aus opportunistischem Verhalten als aus vertrauenswürdigem Verhalten hat. Andernfalls besteht kein Anreiz, sich nicht vertrauenswürdig zu verhalten. Die Sanktion opportunistischen Verhaltens kann beispielsweise in einem verschlechterten Ruf liegen, der zukünftige Transaktionen erschwert. Dies greift jedoch nur, wenn das Verhalten tatsächlich in einer Verschlechterung des Rufs resultiert – Voraussetzung hierfür ist, dass andere Personen vom opportunistischen Verhalten erfahren. Eine Möglichkeit, dies bei Anwendungen in Rechnernetzen zu erreichen, sind Reputationssysteme, die in der Folge diskutiert werden.

Sind alle vier Voraussetzungen erfüllt, so ist also schon aus ökonomischer Sicht ein Schutz des Vertrauens durch das Recht wünschenswert; daneben tritt als Motivation die allgemeine positive Einschätzung des Vertrauens in der Gesellschaft (vgl. Abschnitt 4.1.1.1).

4.3 Vertrauenserzeugung in Netzen

Für das Verständnis der Vertrauenserzeugung in Peer-to-Peer-Netzen ist es hilfreich, analoge Mechanismen, wie sie außerhalb von Rechnernetzen angewendet werden, zu betrachten. Im Folgenden wird dann die Übertragung dieser Mechanismen auf Rechnernetze, insbesondere die Anwendung von Reputationssystemen in solchen Netzen, betrachtet.

4.3.1 Klassische Reputationsmechanismen

Reputationssysteme haben im Zusammenhang mit der Verbreitung des E-Commerce in der Literatur seit einigen Jahren Aufmerksamkeit gefunden. Doch schon vorher bestand die Notwendigkeit, sich über die Vertrauenswürdigkeit anderer Personen oder Institutionen auszutauschen, und entsprechende Mechanismen sind seit langer Zeit etabliert.

Erbringt z.B. ein Dienstleister in einer Kleinstadt Dienstleistungen mangelhafter Qualität, so wird sich dies bald herumsprechen. Auch Bewohner der Stadt, die keine eigenen Erfahrungen mit diesem Dienstleister gemacht haben, erfahren von diesen Problemen und verbreiten die Information vielleicht sogar selbst weiter. Problematisch ist allerdings die oft mangelnde Überprüfung der behaupteten Tatsachen. So könnte der Ausgangspunkt entsprechender Berichte auch ein Konkurrent des Dienstleisters sein, der auf diese Art und Weise Kunden abwerben möchte.

Wer nun entscheiden möchte, ob er einer Information glauben schenken soll, der kann dazu verschiedene Aspekte in seine Überlegungen einbeziehen:

- Aufgrund eigener Erfahrung lässt sich eine subjektive, von der Informationsquelle unabhängige Wahrscheinlichkeit bestimmen, mit der die Information den Tatsachen entspricht.
- Die Vertrauenswürdigkeit der Informationsquelle kann bewertet werden.
- Gibt die Informationsquelle an, von wem die Information stammt, so lässt sich unter Umständen auch die Vertrauenswürdigkeit dieser Person bewerten.
- Zusätzlich kann miteinbezogen werden, wie viele Zwischenschritte zwischen Erfahrung aus erster Hand und eigener Kenntnisnahme der Information zurückgelegt wurden.

Voraussetzung, um diese Methoden anwenden zu können, ist die Fähigkeit, die Vertrauenswürdigkeit einer Person einzuschätzen. Dazu verwendete Informationen können sein:

- Eigene, frühere Erfahrungen mit dieser Person. Hat die Person sich früher vertrauenswürdig verhalten?
- Die Erfahrungen anderer Personen, insbesondere solcher, denen man selbst vertraut, mit der zu beurteilenden Person.
- Wenn beides nicht möglich ist, ist nur noch der Rückgriff auf den persönlichen Eindruck möglich. Dieser kann z.B. durch die Kleidung der Person, ihre Stimme oder ihr Auftreten geprägt sein.

Der persönliche Eindruck, den eine Person hinterlässt, kann in weiten Grenzen von dieser Person direkt beeinflusst werden. Deshalb sind offenbar die ersten beiden Kategorien (eigene Erfahrungen und die Erfahrungen Anderer) vorzuziehen. Sie sind eng verknüpft mit der *Reputation*.

Definition 5. Die Reputation eines Knotens ist seine Vertrauenswürdigkeit (also die wahrgenommene Wahrscheinlichkeit, mit der er sich erwartungsgemäß verhalten wird) in der Wahrnehmung einer Gruppe von Knoten. Sie basiert auf seinem vergangenen Verhalten.

Vertrauensentscheidungen auf Reputation zu stützen, gründet sich

- auf die (empirisch begründete) Annahme, das Verhalten bzw. die Vertrauenswürdigkeit einer Person ändere sich über die Zeit nicht.

- auf den wirtschaftlichen Wert der Reputation: Ein guter Ruf kann zukünftige Transaktionen erleichtern; dies reduziert die Wahrscheinlichkeit, dass er durch böswilliges Verhalten aufs Spiel gesetzt wird (vgl. [RZFK00, S. 46]).

Es zeigt sich, dass die Vertrauensbildung online ganz ähnlich funktioniert. So können frühere eigene Erfahrungen herangezogen werden oder das Aussehen einer graphischen Oberfläche anstelle der Kleidung Grundlage einer Vertrauensentscheidung sein. Der nächste Abschnitt geht detaillierter auf die online anwendbaren Methoden der Vertrauenserzeugung ein.

4.3.2 Vertrauensbildende Faktoren

In jüngerer Zeit hat die Literatur untersucht, wie Vertrauen ohne unmittelbaren persönlichen Kontakt erzeugt werden kann. Der Schwerpunkt liegt dabei im Bereich des Electronic Commerce.

Für klassische E-Commerce-Angebote haben Patton und Jøsang [PaJø03] eine Reihe solche Möglichkeiten identifiziert:

1. Die Gestaltung des Web-Interface. Wie sich diese im Einzelnen auf die wahrgenommene Glaubwürdigkeit und somit auf das Vertrauen in eine Website auswirken kann, ist in [FMLO+01] empirisch untersucht.

2. Datenschutzstrategien und Datenschutzerklärungen. P3P [P3P06] als Standard für die Kommunikation von Datenschutzerklärungen ist hier eine Unterstützung. Die Einhaltung der Datenschutzerklärung kann so jedoch nicht überwacht werden, so dass auch in Hinblick auf Datenschutz Vertrauen nach wie vor notwendig ist.

3. Humanoide. Durch die Darstellung animierter, menschenähnlicher Figuren, die zu Gestik und Mimik in der Lage sind, soll ein persönliches Gespräch nachempfunden werden. [PaJø03, S. 83] Die Aufführung von Humanoiden als eigener Kategorie erscheint jedoch zweifelhaft; sie sind Teil des Nutzer-Interfaces und können daher mit dem Web-Interface zusammengefasst werden, auch, wenn sie ein besonders hervorstechendes Beispiel darstellen.

4. Selbstregulierung und Gütezeichen. Die Anbieter von Gütezeichen bestätigen, dass die zertifizierte Website bestimmten Standards genügt.

5. Sicherheitsstrategien und ihre Wahrnehmung. Neben der tatsächlichen Absicherung verwendeter Systeme spielt die Darstellung dieser Absicherung nach außen eine große Rolle.

6. Zahlungsintermediäre und Versicherungsdienstleister. Zahlungsintermediäre (wie Kreditkartenanbieter) sind meist in der Lage, die Identität eines Website-Anbieters

zu verifizieren. Außerdem können Zahlungsintermediäre und Versicherungsanbieter das Risiko eines Käufers senken, sofern sie selbst vertrauenswürdig sind. Vertrauen in einen Händler kann also (in Grenzen) durch Vertrauen in den Intermediär ersetzt werden.

7. Das Angebot alternativer Streitschlichtung. Die Streitschlichtung außerhalb der Gerichtsbarkeit kann die Kosten im Streitfall reduzieren.

8. Mathematische Vertrauensmodelle, wie das PGP-Web-of-Trust.

9. Reputationssysteme. Auf diese, wie auch auf mathematische Vertrauensmodelle, wird in Abschnitt 4.3.4 näher eingegangen.

Diese Methoden und Systeme lassen sich in drei Kategorien zusammenfassen:

- Durch den Vertrauensnehmer angewandte Methoden und Systeme, die hauptsächlich zu einer Verringerung der wahrgenommenen Wahrscheinlichkeit führen, dass das Vertrauen des Kunden enttäuscht wird (1, 2, 3, 4, 5).
- Methoden, die hauptsächlich den Schaden des Kunden im Fall missbrauchten Vertrauens reduzieren sollen (6, 7).
- Methoden, die ohne direkte Einwirkungsmöglichkeit des Vertrauensnehmers die Einschätzung seiner Vertrauenswürdigkeit ermöglichen sollen (8, 9).

Die ersten beiden Kategorien betreffen keine spezifischen Fragestellungen des Vertrauens in Netzen und werden daher an dieser Stelle nicht weiter verfolgt. Vor einer näheren Betrachtung der letzten Kategorie soll nun das Konzept transitiven Vertrauens eingeführt werden.

4.3.3 Transitives Vertrauen

Transitives Vertrauen kommt dann zum Einsatz, wenn die Vertrauenswürdigkeit eines Systems oder einer Person beurteilt werden soll, der Beurteilende dies aber nicht selbst tun kann. Er bedient sich dann einer weiteren Person (oder eines Systems), der er vertraut. Transitives Vertrauen liegt vor, wenn der Beurteilende aufgrund der Tatsache, dass die zweite Person dem zu Beurteilenden vertraut, selbst auch Vertrauen in den zu Beurteilenden entwickelt. Wenn Interaktionen mit vielen anderen Personen oder Systemen vorkommen – wie dies beispielsweise in Peer-to-Peer-Netzen üblich ist – ist transitives Vertrauen hilfreich, um überhaupt erst Aussagen über die Vertrauenswürdigkeit eines Großteils der Netzteilnehmer treffen zu können.

Abdul-Rahman und Hailes [ARHa97, S. 50] haben Voraussetzungen für die Transitivität von Vertrauen (die sie „bedingte Transitivität von Vertrauen“ nennen) identifiziert. Demnach folgt aus den Aussagen „A vertraut B“ und „B vertraut C“ die Aussage „A vertraut C“ unter folgenden Voraussetzungen:

- B teilt A ausdrücklich mit, dass er C vertraut.
- A vertraut B als Vertrauenseinführer.
- A darf die Qualität von Bs Empfehlung für C beurteilen.
- Das Vertrauen von A in C kann geringer sein als das von B in C.

Transitives Vertrauen kann implizit verwendet oder explizit modelliert werden. Die implizite Verwendung findet sich oft in Reputationssystemen, die im folgenden Abschnitt eingeführt werden.

4.3.4 Reputation und Reputationssysteme

Ziel von Reputationssystemen ist, das Konzept der Reputation auch in rechnernetzbasierten Interaktionen nutzen zu können. Da zwischen den Nutzern der Systeme keine verbale Kommunikation stattfindet, muss ein Reputationssystem auf einem anderen Weg ermöglichen, die Vertrauenswürdigkeit von Nutzern beurteilen zu können. Resnick et al. [RZFK00, S. 46] stellen dazu fest:

> Ein Reputationssystem sammelt, verteilt und aggregiert Rückmeldungen über vergangenes Verhalten seiner Teilnehmer.[8]

Darauf aufbauend lassen sich Reputationssysteme wie folgt definieren:

Definition 6. Reputationssysteme sind Systeme, die Informationen bereitstellen, die dazu verwendet werden, die Vertrauenswürdigkeit (oder Reputation) von Personen oder Systemen zu beurteilen.

Reputation dient dabei als stellvertretende Variable für die Vertrauenswürdigkeit, die sich in den meisten Fällen nicht direkt beobachten lässt. Der Nutzer ist also an der Vertrauenswürdigkeit eines anderen Nutzers interessiert, stützt sein Urteil aber auf dessen Reputation.

4.3.4.1 Allgemeines

In der Literatur wurden folgende Mindestanforderungen an Reputationssysteme identifiziert:

- Sowohl Nutzer, die vertrauen, als auch solche, denen vertraut wird, sollten langlebige Identitäten haben (vgl. [RZFK00, S. 47]). Somit wird sichergestellt, dass sich die Vornahme einer Bewertung überhaupt lohnt. Was unter Langlebigkeit verstanden wird, ist anwendungsabhängig; zumindest sollte die Lebensdauer so gewählt werden, dass während dieses Zeitraums zahlreiche Transaktionen stattfinden – typischerweise im Bereich von Monaten oder Jahren.

[8] „A reputation system collects, distributes, and aggregates feedback about participants' past behavior."

- In engem Zusammenhang damit steht, dass die Identität eines Teilnehmers entweder mit seinem echten Namen verknüpft werden oder zumindest sichergestellt werden muss, dass ein Nutzer unter einer neu geschaffenen Identität kein Vertrauen genießt (vgl. [KaSGM03, S. 640]). Im ersten Fall wird verhindert, dass ein Angreifer zahlreiche virtuelle Identitäten schafft (Sybil-Attack [Douc02]); im zweiten Fall kann ein Angreifer diese zwar schaffen, hat aber keinen Vorteil davon.
- Informationen sollten so zur Reputation eines bewerteten Teilnehmers aggregiert werden, dass sie für den Nutzer verständlich sind und als Grundlage für eine Vertrauensentscheidung dienen können.

Als Beispiel kann das eBay-Reputationssystem [ReZe02] dienen: Die Verknüpfung der Identität eines Teilnehmers mit dessen realem Namen und Adresse (und somit auch eine gewisse Langlebigkeit – Identitäten bleiben typischerweise über mehrerere Jahre hinweg gültig) wird durch den Plattformbetreiber hergestellt. Nach jeder Transaktion können beide Transaktionspartner jeweils eine Bewertung abgeben, die positiv, neutral oder negativ sein und zusätzlich eine textuelle Information enthalten kann. Die Bewertungspunktzahl – also der aggregierte Reputationswert – eines Teilnehmers ergibt sich als Differenz der Anzahl positiver und negativer Bewertungen; außerdem werden einige zusätzliche statistische Maße zur Verfügung gestellt.

Trotz offensichtlicher Schwächen – so beispielsweise der fehlenden Gewichtung einzelner Transaktionen, die dazu führt, dass ein guter Reputationswert durch eine Vielzahl geringwertiger Transaktionen erreicht werden kann, aber für die Beurteilung auch bei anstehenden Geschäften erheblicher Bedeutung herangezogen wird – kann dieses System als Grundlage für eine Vertrauensentscheidung dienen. Insbesondere ist der aggregierte Reputationswert leicht zu interpretieren, da sein Zustandekommen auch für Laien verständlich ist.

Das eBay-Reputationssystem realisiert ein zentralisiertes Reputationsmanagement. Das bedeutet, dass die Reputationswerte der Teilnehmer werden an einer zentralen Stelle errechnet und direkt von dort abgefragt werden.

4.3.4.2 Verteilte Reputationssysteme

Während der Betrieb eines Reputationssystems auf einem zentralen Server gut beherrschbar ist, stellen verteilte Reputationssysteme eine größere technische Herausforderung dar. Die Forschung konzentriert sich derzeit auf Peer-to-Peer-Systeme als wichtigstes Anwendungsszenario [JøIB07]. Da diese derzeit überwiegend für den Zweck des Austauschs von Dateien (oft mit multimedialen Inhalten), also Filesharing, genutzt werden, liegt der Anwendungsbereich der verteilten Reputationssysteme meist darin, Netzteilnehmern eine schlechte Reputation zuzuweisen, die gar keine Dateien, nur solche mangelhafter Qualität oder Viren, Würmer und trojanische Pferde verbreiten.

Beispiele sind das Credence-Reputationssystem [WaSi05] und der EigenTrust-

Algorithmus [KaSGM03], der ebenfalls als Grundlage eines Reputationssystems dienen kann. Der EigenTrust-Algorithmus berechnet einen globalen Reputationswert für jeden Netzteilnehmer, das heißt, sein Reputationswert ist aus Sicht aller anderen Knoten gleich. Dazu propagieren Knoten im Peer-to-Peer-System ihre Vertrauenseinstufungen – also Werte, die aussagen, für wie vertrauenswürdig sie andere Knoten halten. Jeder Knoten gewichtet die Vertrauenseinstufungen, die er von anderen Knoten empfängt, mit seiner eigenen Einschätzung von deren Vertrauenswürdigkeit. Indem jeder Netzteilnehmer diesen Schritt mehrfach durchführt, konvergieren die Reputationswerte, die über jeden Teilnehmer gespeichert sind, gegen seinen globalen Reputationswert. Die Speicherung des Reputationswerts eines Knotens erfolgt jeweils bei mehreren anderen Knoten, um sicherzustellen, dass keine Manipulationen möglich sind.

Schwachpunkte sind bislang bei vielen Reputationssystemen

- das Fehlen einer expliziten Modellierung der Domäne, auf die sich das Reputationssystem bezieht. So kann ein Reputationssystem beispielsweise ein Urteil über die Zuverlässigkeit einer Person als Geschäftspartner liefern – es ist aber nicht möglich, automatisiert zu entscheiden, ob diese Person im Kontext einer bestimmten Interaktion als vertrauenswürdig gelten kann: Selbst wenn die Domäne der Interaktion bekannt ist, kann sie nicht mit der Domäne des Reputationswerts abgeglichen werden.
- der Versuch vieler Systeme, einen globalen Vertrauenswert zu berechnen – auch, wenn in vielen Fällen eine lokale Sichtweise sinnvoller wäre.
- das Fehlen einer expliziten Modellierung transitiven Vertrauens. Diese ist nötig, um automatisiert Vertrauensentscheidungen treffen zu können, wenn direktes Vertrauen nicht vorliegt.

Mathematische Vertrauensmodelle können einen Teil dieser Schwächen umgehen.

4.3.4.3 Mathematische Vertrauensmodelle

Mathematische Vertrauensmodelle, in [PaJø03] als Alternative zu Reputationssystemen aufgeführt, sind hier bewusst als Spezialfall aufgeführt: Auch sie stellen Informationen bereit, die dazu dienen können, die Vertrauenswürdigkeit von Vertrauensnehmern zu beurteilen, und fallen somit unter die Definition von Reputationssystemen. Dabei wird jedoch nicht versucht, einen einzelnen, globalen Reputationswert zu berechnen; vielmehr wird eine lokale Sicht auf Vertrauensbeziehungen bereitgestellt.

Als Beispiel für mathematische Vertrauensmodelle wird in der Literatur oft das PGP-Web-of-Trust [Abdu97] herangezogen. Sein Ziel ist, Vertrauen in die Authentizität eines öffentlichen Schlüssels herzustellen, der zur Verschlüsselung oder zur Verifikation digitaler Signaturen dienen kann. PGP verfolgt dabei einen dezentralen Ansatz: Grundsätzlich kann jeder

Teilnehmer des Web-of-Trust den öffentlichen Schlüssel jedes anderen Teilnehmers signieren und damit die Zuordnung des Schlüssels zu einer Person bestätigen. Dabei ist jedoch nicht jede Signatur gleich viel wert, denn nicht jedem wird die gleiche Befähigung zugetraut, die Identität Anderer zu überprüfen.

Zunächst wird in PGP also nur eine Domäne modelliert, in der Vertrauen hergestellt wird: Es handelt sich um das Vertrauen in die Authentizität bestimmter öffentlicher Schlüssel. Hinzu kommt jedoch eine zweite Domäne, nämlich das Vertrauen in die Fähigkeit eines Schlüsselinhabers, als Vertrauenseinführer zu fungieren (also das Vertrauen in die Authentizität der Schlüssel anderer Personen herzustellen). Das OpenPGP-Vertrauensmodell [RFC2440] sieht sogar transitives Vertrauen vor, das zwischen Teilnehmern kommuniziert werden kann; in der Praxis wird dies jedoch kaum verwendet und stattdessen das Vertrauen in Vertrauenseinführer lediglich lokal verwaltet.

Trotz der Beschränkung auf den Bereich der Schlüsselverwaltung kann PGP als Vorbild für andere Reputationssysteme dienen. Explizite Modellierung transitiven Vertrauens und die Berechnung lokaler Sichten statt globaler Reputationswerte könnten für weitere Einsatzbereiche interessant werden.

Reputationssysteme sind jedoch nicht nur aus Sicht von Soziologie und Informatik von Interesse; durch die Beliebtheit des eBay-Reputationssystems [ReZe02] hat außerdem auch das Recht der Reputationssysteme in jüngerer Zeit an Bedeutung gewonnen. Da negative Bewertungen erheblichen ökonomischen Schaden verursachen können, sind sie regelmäßig Gegenstand von Rechtsstreitigkeiten. Dabei stellt sich nicht nur die Frage, unter welchen Umständen eine Bewertung zurückgenommen werden muss, sondern auch die Frage nach der Verantwortlichkeit des Plattformbetreibers. Diese Fragestellungen werden in den folgenden Abschnitten erörtert.

4.3.5 Reputationssysteme und Recht

Wie oben ausgeführt, stellen Reputationssysteme im Wesentlichen Informationen bereit. Es lassen sich zwei Fälle unterscheiden:

- Die Information kann aus (ggf. nachprüfbaren) Tatsachen bestehen; eine solche Tatsache könnte beispielsweise sein, dass der Bewertete eine Rechnung beim Bewertenden nicht bezahlt hat. Begünstigt wird diese Form der Informationsbereitstellung, wenn die Abgabe einer Bewertung an das Stattfinden einer Transaktion geknüpft ist.
- Die Information kann aus subjektiven Einschätzungen bestehen, die die Vertrauenswürdigkeit der Bewerteten betreffen.

Im Folgenden wird gezeigt, wie beide Fälle im deutschen Recht behandelt werden.

4.3.5.1 Rechtmäßigkeit von Bewertungen

Ein Anspruch auf Unterlassung bzw. Schadensersatz könnte sich bei Bewertungen ergeben aus (vgl. [ScLa05, S. 208])

- §§ 1004, 823 Abs. 2 BGB. Als Schutzgesetze kommen §§ 185 (Beleidigung), 186 (üble Nachrede) und 187 (Verleumdung) StGB in Frage. Dies wird jedoch wohl nur in Ausnahmefällen relevant [ScLa05, S. 208].
- § 824 BGB (Kreditgefährdung). Dies setzt die Verbreitung einer nicht wahrheitsgemäßen Tatsachenbehauptung voraus.
- § 826 BGB (Vorsätzliche sittenwidrige Schädigung). Die Voraussetzungen dieser Vorschrift sind recht hoch, so dass sie nur in Ausnahmefällen Anwendung finden kann. Gefordert ist eine „besondere Verwerflichkeit" des Verhaltens des Schädigers, die sich insbesondere aus deren Zweck oder den eingesetzten Mitteln ergeben kann (Sprau in [Pala06, Rn. 4 zu § 826 BGB]).
- §§ 1004, 823 Abs. 1 BGB. § 823 Abs. 1 BGB schützt eine Reihe absoluter Rechte; wer vorsätzlich oder fahrlässig eines dieser Rechte verletzt, ist dem Verletzten zum Schadensersatz verpflichtet. Im Fall von Reputationssystemen kommt das allgemeine Persönlichkeitsrecht[9] (aus Artikel 1 Abs. 1 GG und Artikel 2 Abs. 1 GG) oder das Recht am eingerichteten und ausgeübten Gewerbebetrieb des Bewerteten als „sonstiges Recht" im Sinne des § 823 Abs. 1 in Frage. Diese Norm wird in der Regel herangezogen, wenn keine Tatsachenbehauptung, sondern ein Werturteil vorliegt. Das Recht am eingerichteten und ausgeübten Gewerbebetrieb hat allerdings einen eingeschränkten Anwendungsbereich: Ein Gewerbebetrieb soll gegenüber einer natürlichen Person nicht privilegiert werden [vSta99, Rn. D2]. Zudem handelt es sich lediglich um einen Auffangtatbestand, der eine Lücke im geschriebenen Recht schließt – aber nur, soweit diese Lücke auch besteht [vSta99, Rn. D20], Sprau in [Pala06, Rn. 126 zu § 823 BGB]. Speziellere Normen wie § 824 BGB oder wettbewerbesrechtliche Regelungen gehen dem Recht am eingerichteten und ausgeübten Gewerbebetrieb also vor. Ob dieses Recht überhaupt ein „sonstiges Recht" im Sinne des § 823 Abs. 1 BGB ist, ist zwar umstritten, doch ist dies nur von beschränkter praktischer Relevanz [vSta99, Rn. D3]. Dörre/Kochmann [DöKo07, S. 33] weisen zudem darauf hin, dass eine einzelne Bewertung nur in den seltensten Fällen einen Gewerbebetrieb in seiner Gesamtheit beeinträchtigen kann, was aber Voraussetzung für die Verletzung dieses Rechts wäre.
- §§ 280 Abs. 1, 241 Abs. 2 BGB im Fall bestehender vertraglicher Beziehungen. § 241 Abs. 2 BGB regelt Nebenpflichten aus einem Schuldverhältnis, zu denen die Rücksichtnahme auf Rechte und Interessen der anderen Partei gehört. Dörre/Kochmann [DöKo07, S. 36 f.] leiten eine solche Nebenpflicht im Fall des eBay-Reputationssystems aus den Allgemeinen Geschäftsbedingungen des Plattformbetreibers her; es ist jedoch

[9] In der Rechtsprechung beispielsweise in [OLGOL06] herangezogen.

davon auszugehen, dass auch ohne solche AGB eine Pflicht besteht, bei der Formulierung der Bewertung Rücksicht auf den Vertragspartner zu nehmen[10]. Eine solche Pflicht geht dann zwar nicht so weit wie im Beispiel der eBay-AGB – die eine „sachliche", d.h. auf Tatsachen gestützte Bewertung verlangen[11] – gefordert. Sie schließt aber zumindest solche Bewertungen aus, die vorwiegend der Schädigung des Vertragspartners dienen.

- §§ 3, 4 Nr. 8 UWG. Wenn ein Wettbewerber eine ungerechtfertigte Bewertung abgibt, beeinträchtigt dies den Wettbewerb zum Nachteil des Bewerteten. Bei einer mehr als unerheblichen Beeinträchtigung ist diese Handlung unzulässig (§ 3 UWG, konkretisiert durch das Beispiel in § 4 Nr. 8 UWG). Daraus folgt ein Unterlassungs (§ 8 Abs. 1 UWG)- und Schadensersatzanspruch (§ 9 UWG).

In allen Fällen gilt es bei der Auslegung das Spannungsverhältnis grundrechtlich geschützter Rechtsgüter zu beachten: Auf der einen Seite das Recht auf freie Meinungsäußerung (Art. 5 I Satz 1 GG) des Bewertenden, auf der anderen das Allgemeine Persönlichkeitsrecht bzw. das Recht am eingerichteten und ausgeübten Gewerbebetrieb (oder sonstige, z.B. durch Schutzgesetze im Sinne des § 823 Abs. 2 BGB geschützte Rechtsgüter) des Bewerteten.

Unter einer Meinung, die durch Art. 5 I GG geschützt ist, versteht man jede Äußerung, die „durch die Elemente der Stellungnahme, des Dafürhaltens oder Meinens geprägt ist" [BVerfG82]. Somit wird auch deutlich, wieso eine Unterscheidung zwischen durch das Reputationssystem bereitgestellten Tatsachen und subjektiven Einschätzungen zu treffen ist: Während subjektive Einschätzungen einer Person grundrechtlich geschützt sind, gilt das für reine Tatsachenbehauptungen nicht. Die Abgrenzung ist jedoch oft schwierig. Wahre Tatsachenbehauptungen sind stets unproblematisch und müssen daher an dieser Stelle nicht betrachtet werden. Unwahre Tatsachenbehauptungen *als solche* sind nicht grundrechtlich geschützt.[12] Auch eine Äußerung, in der Elemente der Meinungsäußerung mit Tatsachenbehauptungen vermischt werden, kann in den Schutzbereich des Art. 5 I GG fallen. Es gilt jedoch im Einzelfall abzuwägen, ob eine Äußerung vorwiegend Tatsachenbehauptung ist, oder ob die Nennung zur Unterstützung der Meinungsbildung dienen soll (vgl. dazu [BVerfG82]), es sich also um „meinungserhebliche Tatsachenbehauptungen" [NoTa04, S. 111] handelt.

Selbst wenn eine Äußerung als Meinungsäußerung unter den Schutzbereich des Art. 5 I Satz 1 GG fällt, gibt es allerdings Grenzen für ihre Zulässigkeit. Nach Art. 5 II GG findet das Recht auf freie Meinungsäußerung seine Schranken „in den Vorschriften der allgemeinen Gesetze [...] und in dem Recht der persönlichen Ehre". Zwischen dem Recht auf freie Meinungsäußerung und dem anderen zu schützenden Rechtsgut (i.d.R. die Ehre des Bewerte-

[10] Diese Auffassung findet sich auch in der Rechtsprechung, siehe z.B. [AGEr04]

[11] Dörre/Kochmann [DöKo07, S. 37]

[12] „Unrichtige Information ist unter dem Blickwinkel der Meinungsfreiheit kein schützenswertes Gut, weil sie der verfassungsrechtlich vorausgesetzten Aufgabe zutreffender Meinungsbildung nicht dienen kann" [BVerfG80, S. 2073]

ten) ist eine Abwägung vorzunehmen. Lediglich bei Vorliegen einer Schmähkritik erübrigt sich diese Abwägung [BVerfG01, Absatz 19]. Schmähkritik ist dabei diejenige Kritik, die der „Diffamierung, Herabwürdigung und Schädigung des Betroffenen“ [ScLa05, S. 209] dienen soll, wobei die Sache, um die es vorgeblich geht, in den Hintergrund tritt[13]. Bei Bewertungen, denen sich der Bewertete freiwillig unterwirft, sind etwaige Meinungsäußerungen i.d.R. auch *nur* in diesem Fall rechtswidrig: Wer Waren oder Dienstleistungen anbietet oder sich gar bewusst bei einem Bewertungsportal anmeldet, der profitiert auch von den Bewertungsmöglichkeiten und ist daher nur in geringerem Maße schutzwürdig. Die Grenze ist erst bei einem „stark herabsetzenden Inhalt“ [ScLa05, S. 209] überschritten, von dem i.d.R. nur bei der oben erwähnten Schmähkritik ausgegangen wird.

Während die Rechtsprechung sich auf textuelle Bewertungen direkt anwenden lässt, fällt eine Übertragung auf numerische Bewertungen auf einer Ordinal- oder Kardinalskala schwer. So ist trotz diverser Urteile zum eBay-Bewertungssystem [AGPe05, LGD004, LGKN04, AGEr04, OLGOL06] keine Rechtsprechung bekannt, die sich auf eine lediglich „negative“ Bewertung ohne textuelle Erläuterungen bezieht. In aller Regel wird der Nachweis, dass es bei einer Bewertung wie „-1“ lediglich um Diffamierung und Schädigung des Bewerteten geht, schwer zu führen sein. Gänzlich ausgeschlossen ist dies indes nicht. Oft ist zu einer numerischen Bewertungsskala auch eine Art „Übersetzungstabelle“ angegeben, die Anhaltspunkte für die Wahl eines Zahlenwertes geben kann. Ist beispielsweise eine Transaktion reibungslos abgelaufen, so soll ein positiver Wert vergeben werden. Wenn nun trotz reibungslosen Ablaufs ein negativer Wert vergeben wird, so ist dies ein Indiz für die Diffamierungsabsicht (die wiederum zum Vorliegen einer rechtswidrigen Schmähkritik führen kann) – oder einen schlichten Irrtum.

Auch die Person des Bewertenden kann ein solches Indiz sein. Gibt beispielsweise ein Wettbewerber des Bewerteten eine ungewöhnlich negative Bewertung ab, so liegt nahe, dass dieser in Wettbewerbsabsicht handelt und auf eine Herabsetzung des Bewerteten zielt.

Es gilt jedoch nicht nur zu beachten, auf welcher Skala bewertet wird – eine größere Rolle spielt, *was* bewertet wird. Je subjektiver Werturteile im jeweiligen Bereich sind, desto mehr entziehen sie sich der richterlichen Inhaltskontrolle. Wird die Schnelligkeit der Abwicklung einer konkreten Transaktion bewertet, sind die Bewertungen noch vergleichsweise objektivierbar. Gibt eine Bewertung jedoch an, ob der Bewertende bereit wäre, dem Bewerteten ein Geheimnis anzuvertrauen, so kann dies völlig ohne rationale Grundlage geschehen. In diesem Fall kann eine gerichtliche Korrektur aufgrund des grundrechtlichen Schutzes der Meinungsäußerung nicht erfolgen.

Bei allen Problemen, die durch einzelne negative Bewertungen entstehen, gilt es indes zu bedenken, dass diese zum Teil nur in primitiven Reputationssystemen auftreten. Berück-

[13] In der Rechtsprechung wurden selbst Äußerungen wie „Achtung Betrüger unterwegs“ – abhängig vom Kontext, in dem sie fallen – als zulässige Meinungsäußerungen eingeordnet, die die Grenze der Schmähkritik nicht überschreiten [OLGKo07].

sichtigt der Algorithmus zur Aggregation von Bewertungen wiederum die Reputation des Bewertenden – wie das beispielsweise beim EigenTrust-Verfahren [KaSGM03] der Fall ist – , sinkt die Wahrscheinlichkeit ungerechtfertigter Negativbewertungen erheblich. Andererseits werden durch die Aggregation neue Angriffsmöglichkeiten eröffnet: Prüft der Nutzer nicht mehr einzelne textuelle Bewertungen, sondern verlässt sich auf einen Aggregationsalgorithmus – im einfachsten Fall die Durchschnittsbildung über Bewertungszahlen – , so ist dieser möglicherweise Angriffen ausgesetzt. Neben der Berücksichtigung von Sicherheitserwägungen beim Entwurf dieser Algorithmen könnten auch rechtliche Möglichkeiten bei der Abwehr von Angreifern bestehen. Auch, wenn im vorliegenden Abschnitt Aggregationsalgorithmen für Reputationssysteme diskutiert werden, lassen die Ergebnisse sich doch auf Empfehlungsalgorithmen übertragen, wie sie in Kapitel 5 eingeführt werden.

4.3.5.2 Angriffe auf Aggregationsalgorithmen

Über die rechtliche Bewertung solcher Angriffe findet sich bislang keine Literatur. Deshalb gilt es zunächst herauszuarbeiten, wann die Beeinflussung eines Algorithmus zur Aggregation von Bewertungen rechtswidrig ist. Anschließend wird die Anwendbarkeit straf- und zivilrechtlicher Normen geprüft. Zunächst wird dabei vom Fall einer bestehenden zentralen Plattform ausgegangen. Anschließend wird diskutiert, inwieweit das gefundene Ergebnis für verteilte Mechanismen haltbar ist.

Die Einflussnahme auf Bewertungsalgorithmen ist nicht grundsätzlich rechtswidrig: Schon die Abgabe einer Bewertung stellt eine solche Einflussnahme dar. Welche Einflussnahme noch erlaubt ist, ergibt sich zunächst, soweit der Angreifer in einem Vertragsverhältnis mit dem Plattformbetreiber steht, aus diesem Vertrag.

Vertragliche Regelungen Vertragliche Regelungen erscheinen als natürlicher Weg, unerwünschte Einflussnahme auf verwendete Bewertungsalgorithmen auszuschließen. Sofern nicht ausdrückliche Regelungen (wie beispielsweise in den eBay-AGB [EBAY07], die in § 18 die Verwendung von „Mechanismen [...], die das Funktionieren der eBay-Website stören können“ untersagen) bestehen, so kann sich das Verbot einer solchen Einflussnahme durch Auslegung des Vertrags über die Nutzung des Bewertungssystems ergeben.

Wird die Grenze des (vertraglich) Erlaubten überschritten, so besteht die Möglichkeit strafrechtlicher Sanktionen und deliktischer Haftung.

Strafrechtliche Bewertung. Aus strafrechtlicher Sicht kommt zunächst § 263a StGB in Betracht. Der relevante Absatz 1 lautet:

> Wer in der Absicht, sich oder einem Dritten einen rechtswidrigen Vermögensvorteil zu verschaffen, das Vermögen eines anderen dadurch beschädigt, dass er das Ergebnis eines Datenverarbeitungsvorgangs durch unrichtige Gestaltung

> des Programms, durch Verwendung unrichtiger oder unvollständiger Daten, durch unbefugte Verwendung von Daten oder sonst durch unbefugte Einwirkung auf den Ablauf beeinflusst, wird mit Freiheitsstrafe bis zu fünf Jahren oder mit Geldstrafe bestraft.

Unproblematisch ist, einen Aggregationsalgorithmus als Datenverarbeitungsvorgang zu qualifizieren. Fraglich sind jedoch die weiteren Voraussetzungen:

- Die „unrichtige Gestaltung des Programms“ kommt in der Regel nicht in Frage, wenn lediglich die Eingaben des Algorithmus manipuliert werden können. Denkbar wäre dieses Tatbestandsmerkmal jedoch, wenn die Daten Programmteile enthalten, die durch Ausnutzen von Sicherheitslücken zur Ausführung kommen (z.B. Buffer Overflows).
- Die „Verwendung unrichtiger oder unvollständiger Daten“ bezieht sich auf inhaltlich falsche Daten (vgl. Perron in [ScLS06, Rn. 6 zu § 263a], Ernst in [Erns04a, S. 105, Rz. 300]), nicht jedoch auf Daten, die gezielt in den Ablauf eines Algorithmus eingreifen.
- Die „unbefugte Verwendung von Daten“ umfasst z.B. die Verwendung von Authentifizierungs- oder Autorisierungsinformationen durch Nichtberechtigte (vgl. [ScLS06, Rn. 8 zu § 263a]), kann also bei einem Angriff auf einen Aggregationsalgorithmus zutreffen.
- Eine sonstige „unbefugte Einwirkung auf den Ablauf“ liegt beispielsweise vor, wenn Daten so gestaltet sind, dass der Algorithmus sie anders als vorgesehen verarbeitet. Ob die Einwirkung unbefugt ist, muss durch eine zivilrechtliche Betrachtung geklärt werden; jedoch ist die Möglichkeit bei einem Aggregationsalgorithmus im Rahmen eines Reputationssystems durchaus gegeben.

Dem Wortlaut der Norm nach sind Angriffe auf Aggregationsalgorithmen also erfasst. Zusätzlich wird in der Literatur aufgrund der Nähe zum Tatbestand des Betrugs jedoch vorausgesetzt, dass der Datenverarbeitungsvorgang „vermögensrelevant“ ist (Perron in [ScLS06, Rn. 20 zu § 263a]) bzw. durch diesen Vorgang eine Vermögensdisposition ausgelöst oder auf ihren Inhalt Einfluss genommen wird [LeWi86, S. 659]. Dies ist bei einem Reputationssystem aber nicht der Fall. Zwar kann die Reputation Grundlage für eine Entscheidung sein, die zu einer Vermögensverfügung führt. Doch ist das nur eine mittelbare Folge der Einwirkung auf den Algorithmus, die für eine Strafbarkeit nach § 263a StGB nicht ausreicht (vgl. Perron in [ScLS06, Rn. 21 zu § 263a]). Die Entscheidung selbst wird durch einen Menschen getroffen. Wird im Ergebnis dadurch ein Mensch getäuscht, so ist denkbar, dass dadurch der Tatbestand des Betrugs (§ 263 StGB) verwirklicht wird.

Denkbar wäre auch eine Strafbarkeit nach § 303a StGB (Datenveränderung). Absatz 1 lautet:

> Wer rechtswidrig Daten (§ 202a Abs. 2) löscht, unterdrückt, unbrauchbar

> macht oder verändert, wird mit Freiheitsstrafe bis zu zwei Jahren oder mit Geldstrafe bestraft.

Die Veränderung von Daten schließt dabei auch die Veränderung eines Computerprogramms ein [Erns04a, S. 95, Rz. 269]. Wer also unbefugt den Aggregationsalgorithmus selbst ändert, der macht sich nach § 303a I StGB strafbar. Eine Einwirkung durch geschickte Wahl der Eingabedaten des Algorithmus wird jedoch i.d.R. nicht dem § 303a StGB unterfallen. In diesem Fall kann sich eine Strafbarkeit lediglich aus § 263 StGB (Betrug) ergeben.

Zivilrechtliche Ansprüche Zivilrechtliche Ansprüche gegen denjenigen, der auf seinen Aggregationsalgorithmus Einfluss nimmt, können sich ergeben aus

- § 826 BGB (vorsätzliche sittenwidrige Schädigung) im Falle von Vorsatz des Täters. Bedingter Vorsatz (der Täter nimmt den Schaden billigend in Kauf) reicht aus [MaEr04, S. 181, Rz. 501].
- § 823 II BGB, wobei die oben aufgeführten strafrechtlichen Normen als Schutzgesetze herangezogen werden können.
- Soweit die Einflussnahme auf den Algorithmus der Diffamierung von Teilnehmern des Reputationssystems dient, so können die gleichen Normen herangezogen werden wie bei einer einzelnen, ungerechtfertigt negativen Bewertung (vgl. S. 102).

In allen Fällen wird sich jedoch der Nachweis eines Schadens und seiner Höhe meist sehr schwierig gestalten.

Fazit In diesem Abschnitt wurde gezeigt, wie das Recht einen Interessenausgleich zwischen Bewertendem und Bewertetem schaffen kann. Dazu wurden Ansprüche gegen den Ersteller einer Bewertung sowie Sanktionen gegen Angreifer auf Bewertungsalgorithmen geprüft. Im Folgenden wird untersucht, welche Ansprüche gegen den Betreiber eines Reputationssystems bestehen könnten.

4.3.5.3 Verantwortlichkeit des Plattformbetreibers

Auch, wenn der Anspruch eines Nutzers auf Entfernung der ihn betreffenden Bewertungen feststeht, stellt sich die Frage, ob er auch einen Anspruch gegen den Plattformbetreiber hat. Fraglich ist einerseits, ob der Plattformbetreiber einerseits zur Auskunft über die Person des Bewertenden verpflichtet ist. Mangels spezialgesetzlicher Regelungen kommt lediglich ein sich aus § 242 BGB (Treu und Glauben) ergebender Auskunftsanspruch in Frage. Voraussetzung dafür wäre das Bestehen eines Hauptanspruchs [BGH89], wie er in einem Anspruch auf Löschung einer Bewertung durchaus vorliegen könnte. Zudem wäre aber auch eine Sonderverbindung zwischen dem Auskunftersuchenden und dem Plattformbetreiber erforderlich. Mit Schmitz/Laun [ScLa05, S. 212] ist davon auszugehen, dass eine solche nicht allein

durch das Vorliegen einer Bewertung über den Auskunftersuchenden im Reputationssystem (in der Literatur auch als Bewertungsplattform bezeichnet) des Plattformbetreibers gegeben ist.

Andererseits ist auch fraglich, ob gegen den Plattformbetreiber selbst ein Anspruch besteht

- auf Löschung bzw. Berichtigung einer Bewertung,
- auf Leistung von Schadensersatz für rechtswidrige Bewertungen

Zur Beantwortung dieser Fragen bedarf es zunächst einer Betrachtung des Telemedienrechts, da sich aus diesem die Verantwortlichkeit eines Plattformbetreibers ergibt. Die Argumentation bezüglich der Einordnung von Empfehlungsdiensten als Telemedien aus Kapitel 3 ändert sich für Reputationssysteme nicht; auch Reputationssysteme sind als Telemedien einzuordnen.

Die Verantwortlichkeit des Plattformbetreibers ist also in den §§ 7-10 TMG geregelt. Nach § 7 I TMG sind Diensteanbieter „für eigene Informationen [...] nach den allgemeinen Gesetzen verantwortlich." Dies gilt auch für „zu eigen gemachte" Informationen[14]. Da die Bewertungen in der Regel nicht vom Betreiber des Reputationssystems stammen, könnte nur der letztgenannte Fall vorliegen.

Eigene Information Zwei Gründe könnten zu der Annahme führen, der Betreiber mache sich die abgegebenen Bewertungen zu eigen: Zum einen genießt der Plattformbetreiber einen wirtschaftlichen Vorteil durch die Bewertungen, zum anderen lässt er sich in vielen Fällen Nutzungsrechte an abgegebenen Bewertungen einräumen. Mit [ScLa05, S. 21] sind beide Argumente jedoch abzulehnen. Ein wirtschaftliches Interesse an der Kommunikation liegt beispielsweise auch bei einem reinen Zugangs-Provider vor, der jedoch nicht für die übermittelte Information haftet. Die Einräumung von Nutzungsrechten andererseits soll lediglich urheberrechtlich begründeten Unterlassungsansprüchen vorbeugen. Für die einzelnen Bewertungen kann also das Vorliegen zu Eigen gemachter Informationen verneint werden.

Anders könnte sich dies jedoch darstellen, wenn der Plattformbetreiber die vorgenommenen Bewertungen aggregiert. Hier lässt sich argumentieren, dass durch die Aggregation keine neue Information entstünde. In der Tat bedeutet Aggregation sogar eine Reduktion vorhandener, von den Nutzern zur Verfügung gestellter Information. Dennoch ist sie aus Anwendersicht und damit auch für die juristische Betrachtungsweise als die Erstellung neuer Information anzusehen. Die Aggregation mag zwar nur dafür gedacht sein, dem Nutzer das Lesen jeder einzelnen Bewertung zu ersparen und ihm trotzdem Gelegenheit zu ge-

[14] Dies geht aus der Gesetzesbegründung der wortgleichen früheren §§ 8–11 TDG [Bund01b, S. 23] hervor und wird so auch in der Literatur [ScLa05, S. 210], [Bedn07, Abs. 24] und Rechtsprechung [OLGK02, OLGD02] vertreten. Entscheidend ist, dass aus Sicht des Nutzers tatsächlich der Eindruck entsteht, dass der Plattformbetreiber sich die Information zu eigen macht.

ben, sich ein Bild des Bewerteten zu machen. Diese Funktion alleine erlaubt zwar erst das Erfassen der vorhandenen Information; sie kann dennoch nicht als das Erzeugen neuer Information bezeichnet werden. Die Aggregation enthält aber immer auch eine Wertung. Je nach Gewichtung unterschiedlicher einfließender Bewertungen kann ein gänzlich anderes Bild eines Bewerteten entstehen. So können Bewertungen

- in Abhängigkeit ihres Alters,
- bei transaktionsbasierten Systemen in Abhängigkeit des Transaktionswerts,
- nach Reputation der Bewerter

und nach weiteren Kriterien gewichtet werden. Es besteht kein technischer Sachzwang, der eine bestimmte Gewichtung zwingend erforderlich machen würde. Da der Plattformbetreiber die Art der Aggregation also in weiten Grenzen frei festlegen kann, ist ihm die erzeugte Information auch als eigene Information zuzurechnen. Diese Frage wurde in der Literatur allerdings bislang nicht diskutiert. Praktisch relevant wird sie ohnehin nur, wenn die Art der Aggregation (und nicht etwa die zugrunde liegenden Daten) angegriffen werden.

Folgerungen Soweit angebotene Informationen als eigene Informationen qualifiziert werden können, so ist der Plattformbetreiber (als Anbieter eines Telemediums) nach § 7 I TMG nach den allgemeinen Gesetzen für diese verantwortlich. Dies gilt also insbesondere für die aggregierten Bewertungen.

Für die Fälle, in denen dies nicht zutrifft, stellt sich die Frage, in welche der Kategorien der §§ 8–10 TMG der Betreiber eines Reputationssystems einzuordnen ist:

- § 8 TMG („Durchleitung von Informationen“) befasst sich mit Anbietern, die fremde Informationen „in einem Kommunikationsnetz übermitteln“ oder den Zugang zu ihrer Nutzung vermitteln. Dies ist aber jedenfalls nicht das Kernfeld der Geschäftstätigkeit des Betreibers eines Reputationssystems. Nach § 8 I Nr. 3 TMG greift die Privilegierung dieser Anbieter ohnehin nur, falls sie „die übermittelten Informationen nicht ausgewählt oder verändert haben.“ Diese Voraussetzung dürfte beim Betreiber eines Reputationssystems i.d.R. nicht erfüllt sein.
- Auch § 9 TMG (Zwischenspeicherung zur beschleunigten Übermittlung von Informationen) trifft nicht zu.
- § 10 TMG regelt die Verantwortlichkeit für fremde Informationen, die ein Diensteanbieter für einen Nutzer speichert. Dass es sich bei den einzelnen Bewertungen um fremde Informationen handelt, wurde oben (S. 108) gezeigt. Fraglich ist aber, ob von der Speicherung „für einen Benutzer“ gesprochen werden kann. Die Tatsache, dass die Bewertungen durch *andere* Benutzer genutzt werden können, schließt die Speicherung „für einen Benutzer“ nicht aus: Die Begründung zu § 11 TDG [Bund01b] erwähnt ausdrücklich die Anwendbarkeit auf Hosting-Provider. Das Hauptinteresse an der Speicherung einer Bewertung hat jedoch nicht der Bewertende, sondern diejenigen, die

eine Entscheidung auf die Bewertung stützen wollen, sowie der Plattformbetreiber. Da die Speicherung allerdings durch den Bewertenden ausgelöst wird, kann von einer Speicherung für ihn ausgegangen werden. Dies steht auch im Einklang mit der Intention des Gesetzgebers (vgl. [ScLa05, S. 211]).

Nach § 10 TMG ist der Betreiber eines Reputationssystems somit nicht für die gespeicherten Bewertungen verantwortlich, sofern er keine Kenntnis von der rechtswidrigen Bewertung[15] hat oder nach Erlangung dieser Kenntnis unverzüglich die Bewertung löscht bzw. den Zugang zu ihr sperrt. Nach § 7 II TMG ist der Betreiber auch nicht zur aktiven Suche nach rechtswidrigen Bewertungen verpflichtet; er kann also den Eingang entsprechender Hinweise abwarten.

Es gilt jedoch zu beachten, dass die Haftungsprivilegierung der §§ 8–10 TMG nicht für Unterlassungsansprüche gilt [Stad05, Rn. 26], [BGH04]. Ein Plattformbetreiber kann folglich als Störer auf Unterlassung in Anspruch genommen werden; dies gilt unabhängig vom Verschulden. Vorausgesetzt wird einerseits die „willentliche und adäquat kausale Herbeiführung" einer rechtswidrigen Beeinträchtigung [Libe07, S. 144], andererseits die Verletzung einer Prüfungspflicht [ScLa05, S. 211]. Die Prüfung muss dem Plattformbetreiber aber zugemutet werden können; dies ist in der Regel erst dann der Fall, wenn Rechtsverletzungen bereits bekannt geworden sind und die technische Möglichkeit besteht, vergleichbare Verletzungen zu unterbinden[16] [ScLa05, S. 212]. Ob dies möglich ist, ist eine Frage des Einzelfalls – jedenfalls kann aber festgehalten werden, dass eine Pflicht zur Vorabprüfung nicht besteht.

4.3.6 Übertragbarkeit auf Empfehlungssysteme

In Literatur und Rechtsprechung hat zwar das Recht der Reputationssysteme bereits erhebliche Aufmerksamkeit gefunden – doch wie lassen sich die gefundenen Erkenntnisse auf Empfehlungssysteme übertragen? Während ein Reputationssystem der Bewertung der Vertrauenswürdigkeit von Personen oder Systemen dient, soll ein Empfehlungssystem die Bewertung von Objekten ermöglichen und die Aufmerksamkeit eines Nutzers auf interessante Objekte lenken helfen. Doch sind nach wie vor textuelle Bewertungen möglich, und auch die Rechte anderer Personen (Hersteller oder Händler) können ebenso verletzt werden. Somit ergibt sich als einziger Unterschied, dass das Allgemeine Persönlichkeitsrecht bei der Bewertung von Objekten in geringerem Maße betroffen ist; eine Bewertung für ein Objekt kann zwar mittelbar auch Auswirkungen auf eine Person haben, doch wird die Abwägung hier im Zweifelsfall eher zu Gunsten des Bewertenden ausfallen als bei der Bewertung einer Person.

[15] Kenntnis der Bewertung reicht aus; es ist nicht gefordert, dass der Betreiber auch weiß, dass diese Bewertung rechtswidrig ist.

[16] Für Auktionsplattformen wurde dies im Fall von Marken- sowie Urheberrechtsverletzungen in der Rechtsprechung aber schon verschiedentlich bejaht [LGM06, BGH04]

Ansprüche, die sich nicht auf das Allgemeine Persönlichkeitsrecht stützen – insbesondere wettbewerbsrechtliche Ansprüche – werden dadurch nicht berührt; hier lässt sich die Argumentation, die für Reputationssysteme herangezogen wurde, direkt auf Empfehlungssysteme übertragen.

4.3.7 Fazit

In diesem Abschnitt wurden Reputationssysteme als ein Mittel zur Erzeugung von Vertrauen vorgestellt und aus juristischer Sicht diskutiert. Im Folgenden soll nun betrachtet werden, wie sich Vertrauen und Vertraulichkeit zueinander verhalten und wie Vertrauen zum Schutz vertraulicher Information eingesetzt werden kann.

4.4 Vertrauen als Konzept zum Schutz von Vertraulichkeit

Vertrauen und Vertraulichkeit sind eng miteinander verknüpft: Wer jemandem eine vertrauliche Information mitteilt, der vertraut ihm diese an. Anders formuliert: Er begibt sich bewusst in eine Abhängigkeit vom Verhalten des anderen, da dieser die Information preisgeben könnte. Dieser Sachverhalt unterfällt also der Definition von Vertrauen (S. 86); wer vertrauenswürdige Informationen teilt, muss dem Empfänger vertrauen.

Dies lässt sich auch auf Empfehlungssysteme anwenden, in denen potentiell vertrauliche Daten verwendet werden können. So kann die Tatsache, dass eine Person ein bestimmtes Medikament empfiehlt, zwar anderen Personen bei ihrer Wahl helfen, aber auch Informationen über den Gesundheitszustand der empfehlenden Person beinhalten. Sind diese Informationen vertraulich, sollten sie nur an vertrauenswürdige Personen weitergegeben werden. Es besteht also ein Bedürfnis zum Schutz bestimmter Informationen, die aber andererseits nicht völlig geheimgehalten werden müssen.

4.4.1 Grundlagen

Vertrauliche Informationen können auf verschiedene Arten geschützt werden:

- Durch das Recht. Hierzu zählen z.B. das Datenschutzrecht (s. Kapitel 3), die ärztliche Schweigepflicht (§ 203 StGB), der wettbewerbsrechtliche Schutz von Geschäfts- und Betriebsgeheimnissen (§ 17 UWG), der Schutz des Fernmeldegeheimnisses (Art. 10 GG, § 85 TKG) usw.
- Durch Zugriffskontrollmechanismen (vgl. z.B. [SCFY96]).
- Durch kryptographische Verfahren, insbesondere Verschlüsselung von Daten.

Diese Schutzmechanismen schließen sich nicht gegenseitig aus, sondern können einander ergänzen. Sie sind hilfreich dabei, Nichtberechtigte vom Zugriff auf Informationen oder

Ressourcen auszuschließen, doch helfen sie nur eingeschränkt bei der Feststellung, wer zum Zugriff berechtigt sein soll.[17]

Da es sich hierbei um eine Vertrauensentscheidung handelt, liegt nahe, sie ähnlich anderen Vertrauensentscheidungen durch ein Reputationssystem zu unterstützen. Jedoch ist der Anwendungsbereich beschränkt, da die Preisgabe von Daten aus dem engsten Persönlichkeitsbereich oftmals nur an enge Vertraute gewünscht wird.

Im nächsten Abschnitt wird ein Verfahren vorgestellt, dass unter Berücksichtigung dieser Problematik eine möglichst gute Unterstützung des Entscheidungsprozesses zu erreichen sucht.

4.4.2 Schutz von Vertraulichkeit auf Basis von Reputation

Grundidee des Verfahrens (vgl. dazu auch [SoZi06]) ist, in einem Peer-to-Peer-System vertrauliche Informationen bereitzustellen. Diese sollen aber nicht allen anderen Knoten zur Verfügung gestellt werden, sondern nur solchen, zu denen der Absender eine (direkte oder indirekte) Vertrauensbeziehung hat. Eine direkte Vertrauensbeziehung besteht dabei, wenn ein Nutzer einem anderen Nutzer aufgrund eigener Erfahrungen einen positiven Vertrauenswert gegeben hat. Die Entstehung direkter Vertrauensbeziehungen wird dabei nicht betrachtet; indirekte Vertrauensbeziehungen werden aus den direkten Vertrauensbeziehungen durch ein Vertrauensmodell hergestellt. Zur Herstellung transitiven Vertrauens werden – ähnlich dem PGP-Web-of-Trust – Vertrauenseinführer eingesetzt. Ein Vertrauenseinführer ist dabei ein Knoten, dem ein anderer Knoten in Bezug auf die Aufgabe vertraut, die Vertrauenswürdigkeit weiterer Knoten einzuschätzen[18].

Klassisch wird die Anforderung der *Vertraulichkeit* so verstanden, dass die Vertraulichkeit einer Nachricht geschützt ist, wenn außer Absender und beabsichtigtem Empfänger niemand ihren Inhalt zur Kenntnis nehmen kann. Eine Nachricht kann also vertraulich (vor unberechtigtem Lesen geschützt) sein oder nicht. Als Verfeinerung wird die Vertraulichkeit nicht binär, sondern in Abstufungen definiert – je höher die Vertraulichkeit, desto kleiner der Kreis derjenigen, die die Nachricht lesen können. In Abstufungen wird auch das Vertrauen in andere Knoten definiert. Zudem wird die Domänenabhängigkeit von Vertrauen berücksichtigt: Der Nutzer ordnet jedes zu verbreitende Dokument einer Domäne zu, und auch Vertrauensbeziehungen sind jeweils nur für eine Domäne gültig. Das vorgestellte Verfahren implementiert auf diese Weise ein einfaches Vertraulichkeitsmodell.

Der Anwendungsbereich ist nicht auf Empfehlungssysteme beschränkt; für diese bietet sich das Verfahren aber an: So kann ein Nutzer Bewertungen von Objekten[19] in einem Peer-

[17] Rollenbasierte Zugriffskontrolle bietet lediglich insofern eine Erleichterung, als Zugriffsberechtigungen nicht mehr an eine konkrete Person, sondern eine Rolle anknüpfen. Die Zuordnung von Personen zu Rollen ist jedoch von Hand vorzunehmen.

[18] Zu den Begriffen Vertrauen und Vertrauenswürdigkeit siehe auch Seite 86

[19] In der Darstellung des Verfahrens werden die Bewertungen allgemeiner als Dokumente bezeichnet.

to-Peer-System an vertrauenswürdige andere Nutzer verbreiten und somit diesen die Berechnung von Empfehlungen ermöglichen. Gleichzeitig kann er jedoch sicherstellen, dass nicht vertrauenswürdige Nutzer nicht erfahren, dass er bestimmte Objekte – beispielsweise die bereits angeführten Medikamente – bewertet hat.

Die folgende Terminologie wird zur Beschreibung des Verfahrens verwendet:

$T_a^D(b,c)$ Vertrauen von Knoten b in Knoten c, bezogen auf Domäne D, aus Sicht von Knoten a.

$I^D(b,c)$ Vertrauen von Knoten b in Knoten c als Vertrauenseinführer, d.h. der Grad, zu dem Knoten b Knoten c vertraut, die Vertrauenswürdigkeit Anderer zu beurteilen. Auch dieser Wert bezieht sich auf Domäne D.

$\mathrm{Conf}(Z)$ Vertraulichkeitseinstufung des Dokuments Z aus Sicht seines Autors. Jedes Dokument gehört zu genau einer bestimmten Domäne D und stammt von genau einem Autor (beide Einschränkungen können aufgehoben werden; darauf wird jedoch im Sinne einer einfacheren Darstellung des Verfahrens im Folgenden verzichtet).

U Menge aller Knoten, die an dem System teilnehmen.

V_i^D Menge der Knoten j, für die gilt: $T_i^D(i,j) > 0$

$E_b(X)$ Inhalt X, mit dem öffentlichen Schlüssel von b verschlüsselt.

$S_a(X)$ Inhalt X, durch a mit einem Zeitstempel versehen und signiert.

Alle Werte T, I und Conf müssen im Intervall $[0;1]$ liegen. Die Messbarkeit von Vertrauen und Vertraulichkeit auf der gleichen Skala wird dabei vorausgesetzt, denn jedes Mal, wenn ein Knoten ein Dokument an einen anderen weiterleitet, prüft er, ob die Vertrauenseinstufung des Empfängers durch den Absender höher ist als die Vertraulichkeitseinstufung des Dokuments (jeweils auf eine Domäne bezogen).

Die Zuweisung von Vertrauenseinführer-Bewertungen in Abhängigkeit von der jeweiligen Domäne wird als notwendig angesehen: Es ist durchaus denkbar, dass eine Person in einer Domäne die Vertrauenswürdigkeit anderer Personen gut beurteilen kann, dies aber für andere Domänen nicht gilt. Die Vornahme dieser Einstufung kann für den Nutzer jedoch sehr komplex werden. Deshalb wird als Vereinfachung vorgeschlagen, dass ein Knoten B einem anderen Knoten C eine allgemeine Vertrauenseinführer-Bewertung $I(b,c)$ zuweist. Der domänenspezifische Wert $I^D(b,c)$ wird dann als Produkt aus dieser Bewertung und dem jeweiligen domänenspezifischen Vertrauenswert bestimmt: $I^D(b,c) = I(b,c) \cdot T_b^D(b,c)$. Die Entscheidung wird aber immer lokal getroffen, ein Knoten kann also von diesem vorgeschlagenen Verfahren stets auch abweichen.

An dieser Stelle sei darauf hingewiesen, dass zur Verringerung der Komplexität des Verfahrens die Vertraulichkeit eines Dokuments nicht als domänenabhängig modelliert ist. Daraus folgt, dass eine bestimmte Information nur einer einzigen Domäne zugewiesen werden kann. Selbstverständlich kann in realen Anwendungen eine Information auch mehreren Domänen zugehörig sein. Eine Erweiterung des vorliegend beschriebenen Verfahrens, die diese Zugehörigkeit zu mehreren Domänen abbilden kann, wäre möglich, wird aber nicht verfolgt, um die Komplexität des Verfahrens nicht weiter zu erhöhen.

Durch das Verfahren soll nun ein Knoten, der an einem Peer-to-Peer-System teilnimmt, in die Lage versetzt werden, zu entscheiden, ob ein Dokument an einen bestimmten anderen Knoten weitergeleitet werden darf. Die Entscheidung wird dabei auf Grundlage des Vertrauens des jeweiligen Dokumentautors in den potentiellen Empfänger getroffen. Die Weiterleitung eines Dokuments ist genau dann zulässig, wenn das Vertrauen des Autors in den vorgesehenen Empfänger größer oder gleich der Vertraulichkeitseinstufung des Dokuments ist. Da der Autor in aller Regel nicht ständig verfügbar sein wird, wird die Entscheidung durch andere Knoten ermöglicht, die dazu ein Vertrauensmodell heranziehen. Vertrauensbeziehungen sollen dabei jedoch nur offengelegt werden, sofern dies nötig ist – also die schon bekannten Vertrauensbeziehungen nicht ohnehin die Weiterleitung des Dokuments ermöglichen.

4.4.2.1 Anforderungen

Um die gewünschte Funktionalität zu erreichen und den Einsatz in einem reinen Peer-to-Peer-System zu ermöglichen, lassen sich im Detail die folgenden Anforderungen an das Verfahren identifizieren:

1. Das Verfahren soll ohne zentrale Komponenten auskommen.

2. Ein Dokument darf nur dann an einen Knoten b weitergeleitet werden, wenn (aus der Sicht des weiterleitenden Knotens a) das Vertrauen des Autors c des Dokuments in b (bezogen auf die Domäne des Dokuments Z) größer oder gleich der Vertraulichkeitseinstufung des Dokuments ist: $T_a^D(c, b) \geq \text{Conf}(Z)$.

3. Aus Gründen des Datenschutzes sollten Vertrauensbeziehungen in einen Knoten nur mit Mitwirkung dieses Knotens offengelegt werden können.
   Sowohl in den meisten Reputationssystemen als auch dem PGP-Web-of-Trust sind Vertrauensbeziehungen öffentlich sichtbar. Diese – wenn möglich – zu verstecken, ist jedoch sinnvoll: Es handelt sich um grundsätzlich schützenswerte personenbezogene Daten, und wenn sie nicht aus einem bestehenden Reputationssystem stammen, sondern direkt definiert sind, werden sie auch nicht auf anderem Wege offengelegt.

4. Ein Knoten M kann $T_m^D(b, c)$ nicht unentdeckt höher als von b erlaubt wählen.
   Andernfalls könnte der Angreifer M nicht nur die Dokumente eines Knotens b an einen nicht vertrauenswürdigen Knoten c weiterleiten, sondern zusätzlich andere Knoten – auch solche, die sich gutwillig verhalten – dazu bringen, das Gleiche zu tun.

5. Die Erlangung mehrerer Identitäten gibt einem Angreifer keinen Vorteil.
   Dies bedeutet Immunität gegen den sogenannten Sybil-Angriff (vgl. Abschnitt 2.1.7).

6. Vertrauenseinstufungen können jederzeit geändert werden.
   Vertrauen ändert sich im Laufe der Zeit, insbesondere ist ein Verlust von Vertrauen möglich.

7. Ständige Erreichbarkeit der Netzteilnehmer sollte nicht erforderlich sein.
   Diese Anforderung ist für den Einsatz in Peer-to-Peer-Systemen wichtig, bei denen nicht von einer hohen Verfügbarkeit einzelner Knoten ausgegangen werden kann.

8. Das Verfahren soll auch skalierbar sein; der Kommunikationsaufwand soll im Vergleich zu einem Peer-to-Peer-System ohne den Schutz der Vertraulichkeit höchstens um einen konstanten Faktor steigen.

Bei diesen Anforderungen wurde der Schwerpunkt auf Sicherheit und Datenschutz gelegt; andere wünschenswerte Eigenschaften, wie Nutzerfreundlichkeit und Effizienz, sind nicht im Fokus dieser Arbeit. Die Aufzählung sollte daher nicht als abschließend betrachtet werden.

Um diese Anforderungen erfüllen zu können, müssen zunächst einige Annahmen zugesichert werden; diese werden im nächsten Abschnitt erörtert.

4.4.2.2 Annahmen

Folgende Voraussetzungen müssen erfüllt sein, um die Funktionalität des Verfahrens sicherstellen zu können.

- (A1) Indirekte Vertrauensbeziehungen werden auf Basis direkter Beziehungen geschaffen. Dazu ist transitives Vertrauen – zumindest in einer eingeschränkten Form – notwendig.

- (A2) Direkte Vertrauensbeziehungen müssen bestehen oder etabliert werden, beispielsweise durch Verwendung eines Reputationssystems. Um direkte Vertrauensbeziehungen aufbauen zu können, sind langlebige Identitäten der beteiligten Knoten nötig, die jedoch nicht notwendigerweise mit Identitäten aus der physischen Welt verknüpft sein müssen.

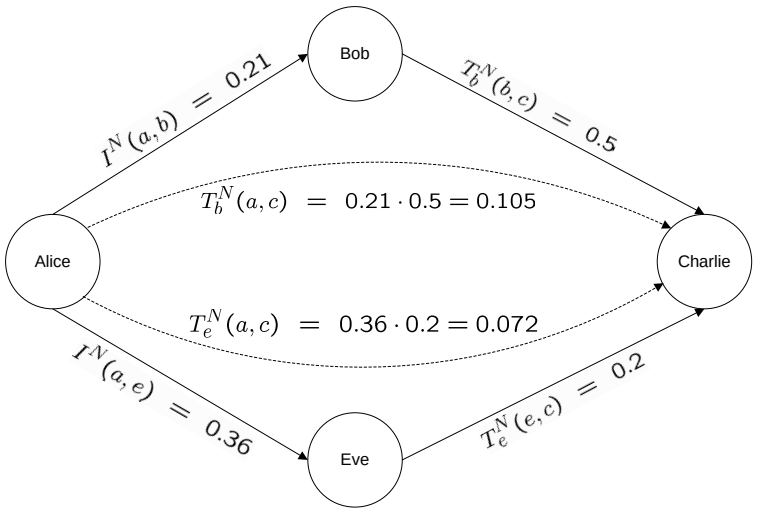

Abbildung 4.2: Ein einfaches Beispiel transitiven Vertrauens

- (A3) Neu hinzukommenden Knoten darf kein positiver direkter Vertrauenswert zugewiesen werden, oder die Erschaffung neuer Identitäten muss schwierig sein. Entweder wird neuen Teilnehmern also nicht vertraut, oder sie müssen eine Leistung erbringen, um Vertrauen zu gewinnen. Mit dieser Annahme wird dem Sybil-Angriff Rechnung getragen.

- (A4) Eine Public-Key-Infrastruktur (PKI) wird benötigt, um die asymmetrische Verschlüsselung von Dokumenten zu ermöglichen. Sie stellt die öffentlichen Schlüssel der Knoten bereit, die das Verfahren anwenden. Die PKI kann durchaus auch bereits die Erfüllung der Annahmen A2 und A3 sicherstellen.

4.4.2.3 Beispiel

In diesem Abschnitt wird das zugrunde liegende Vertrauensmodell anhand eines Beispiels erläutert (vgl. Abbildung 4.2).

Alice gibt Bob in der Domäne „Ernährung“ (N für „nutrition“) einen Vertrauenswert von $0,3$ und eine allgemeine Bewertung als Vertrauenseinführer von $0,7$. Alice berechnet daraus einen domänenspezifischen Wert von $0,21$ für Bob als Vertrauenseinführer:

$$I^N(a,b) \quad = \quad 0,21$$

Bob stuft Charlie in der gleichen Domäne mit einem Vertrauenswert von $0,5$ ein:

$$T_b^N(b,c) \quad = \quad 0,5$$

Aus Bobs Sicht ergibt sich damit:

$$T_b^N(a,c) \quad = \quad I^N(a,b) \cdot T_b^N(b,c) = 0,21 \cdot 0,5 = 0,105$$

Im nächsten Schritt wird Eve von Alice mit einem Vertrauenswert von $0,4$ in der Domäne Ernährung und einer allgemeinen Bewertung als Vertrauenseinführer von $0,9$ eingestuft. Dies führt zu einer domänenspezifischen Bewertung von $0,36$ als Vertrauenseinführer:

$$I^N(a,e) \quad = \quad 0,36$$

Eve vertraut Charlie in der Domäne Ernährung mit einem Wert von $0,2$:

$$T_e^N(e,c) \quad = \quad 0,2$$

Aus Eves Sicht ergibt sich somit:

$$T_e^N(a,c) \quad = \quad I^N(a,e) \cdot T_e^N(e,c) = 0,36 \cdot 0,2 = 0,072$$

Eve könnte ein Dokument aus der entsprechenden Domäne, das von Alice erstellt wurde, also nur dann an Charlie weiterleiten, wenn dessen Vertraulichkeit mit $0,072$ (oder weniger) eingestuft ist. Unter Berücksichtigung von Bobs Einstufung wäre eine Weiterleitung bis zu einer Vertraulichkeit von $0,105$ möglich – dies könnte sogar durch Eve erfolgen, wenn diese Kenntnis von Charlies Einstufung hat.

Aus diesem Beispiel ergibt sich, dass die Existenz direkter Vertrauensbeziehungen zwischen den beteiligten Knoten notwendige Voraussetzung des vorliegenden Verfahrens ist.

Dies gilt sowohl für Vertrauen in der jeweiligen Domäne, andererseits auf das Vertrauen in andere Knoten als Vertrauenseinführer.

Wird ein Dokument an einen anderen Knoten gesandt, werden die Vertrauensbeziehungen des Autors – wie sie im Beispiel dargestellt sind – diesem hinzugefügt. Der Empfänger erfährt somit seine Einstufung I als Vertrauenseinführer. Jegliche weitere Informationen über Vertrauenseinstufungen wird mit dem jeweiligen öffentlichen Schlüssel des bewerteten Knotens verschlüsselt. Der Absender muss also nur diejenigen Vertrauenseinstufungen offenlegen, die tatsächlich benötigt werden, um die Entscheidung über die Weiterleitung eines Dokuments zu treffen; sogar der Empfänger des Dokuments kann die entsprechende Information nicht selbst entschlüsseln.

4.4.2.4 Vertrauensbildung

Als ersten Schritt muss jeder Knoten a die Werte $T_a^D(a,b)$ und $I^D(a,b)$ für mindestens einen Knoten $b \neq a$ bestimmen und lokal speichern. Die Grundlagen für solche Vertrauensentscheidungen können vielfältig sein. Persönliche Beziehungen sind eine Möglichkeit, aber nicht notwendig. Ein Knoten könnte beispielsweise auch positive Vertrauenseinstufungen auf der Basis der IP-Adressen anderer Knoten (beispielsweise im selben Unternehmensnetz), auf Basis des Landes, in dem sie sich befinden, oder des Vorhandenseins von Identitätszertifikaten vertrauter Zertifizierungsstellen vornehmen. Schließlich könnten auch Einstufungen aus einem bereits bestehenden Reputationssystem herangezogen werden. Indirekte Vertrauensbeziehungen werden aus direkten mit Hilfe der folgenden Beziehung berechnet:

$\forall x,y,z$ mit $x,y,z \in U, z \in V_x^D : T_x^D(y,z) = I^D(y,x) \cdot T_x^D(x,z)$, d.h. das Vertrauen von x in z wird diskontiert, da y nur beschränktes Vertrauen in x als Vertrauenseinführer hat. Durch diese multiplikative Verkleinerung wird erreicht, dass nur kurze Vertrauenspfade entstehen – das heißt, das Vertrauen nur über wenige Zwischenknoten aufgebaut wird. So soll transitives Vertrauen realistisch modelliert werden: Auch in der Realität mag man dem Freund eines Freundes vertrauen, nicht aber jemandem, den man nur über zahlreiche Zwischenschritte kennt. Wird mehr als ein Vertrauenspfad gefunden, so wird der höchste der Vertrauenswerte verwendet.

Die Pfadfindung wird dabei durchgeführt, indem der Dijkstra-Algorithmus auf einen Graphen (den *Vertrauensgraphen*) mit den Nutzern als Knoten und ihren Vertrauenswerten als Kanten angewendet wird; im Gegensatz zur gängigen Variante dieses Algorithmus werden die Kantengewichte im betrachteten Vertrauensgraphen nicht addiert, sondern multipliziert. Gesucht wird dann für jeden Zielknoten der Pfad, bei dem das Produkt der Kantengewichte maximal ist.

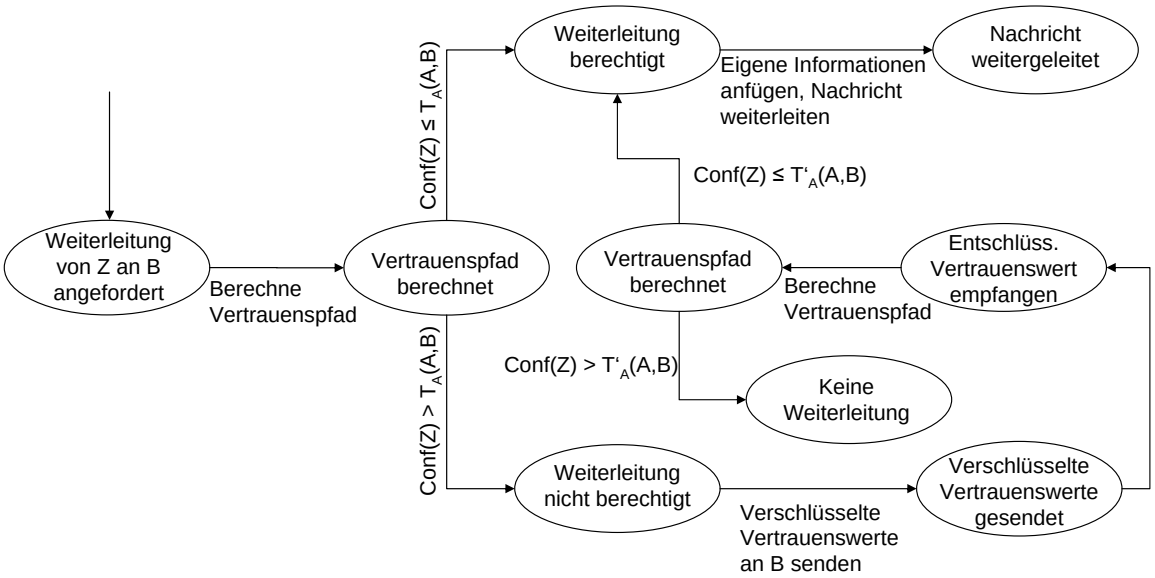

Abbildung 4.3: Übertragung eines vertraulichen Dokuments

4.4.2.5 Übertragung eines Dokuments

Dieser Abschnitt beschreibt das Verfahren, das der Übertragung eines Dokuments zugrunde liegt.

Erstellung eines Dokuments Bei der Erstellung eines Dokuments Z legt der Autor a dessen Vertraulichkeit $\text{Conf}(Z)$ fest, wobei $0 \leq \text{Conf}(Z) \leq 1$. Zudem wählt er eine Domäne D, der das Dokument angehört. Er signiert Dokument und Vertraulichkeitseinstufung: $S_a(Z,\text{Conf}(Z))$.

Übertragung eines Dokuments Das Verfahren zur Übertragung eines Dokuments aus Sicht eines einzelnen Knotens ist in seinen Grundzügen in Abbildung 4.3 dargestellt. Will ein Knoten b ein Dokument Z, das von a erstellt wurde, an einen Knoten f weiterleiten, so muss b prüfen, ob f ausreichend vertrauenswürdig ist, d.h. ob $T_b^D(a,f) \geq \text{Conf}(Z)$. Zunächst wird aufgrund der lokal vorliegenden Information ein Vertrauenspfad von a nach f berechnet. b bestimmt $T_b^D(a,f)$ auf eine der folgenden Arten:

- Der Wert könnte von früheren Übermittlungen zwischengespeichert sein.
- Seien $I^D(a,b)$ und $T_b^D(b,f)$ gegeben, so setzt b: $T_b^D(a,f) = I^D(a,b) \cdot T_b^D(b,f)$.
- Kennt B die Vertrauensbewertungen durch andere Knoten, so kann er den Vertrauenspfad von a nach f wie in Abschnitt 4.4.2.4 beschrieben berechnen.

Ergibt sich, dass das Vertrauen des Autors in den Zielknoten aus Sicht des weiterleitenden Knoten größer (oder gleich) ist als die Vertraulichkeitseinstufung des Dokuments – es gilt also $T_b^D(a, f) \geq \text{Conf}(Z)$ – ,so heißt das, dass die Weiterleitung berechtigt ist. In diesem Fall sendet b an f eine Nachricht M mit folgenden vier Einträgen als Inhalt:

- Falls b der Ersteller des Dokuments ist (also $b = a$): $S_b(Z,\text{Conf}(Z))$
  Sonst: Die Nachricht M', die b von seinem Vorgänger im Pfad erhalten hat.
- $S_b(I^D(b, f))$, d.h. das Vertrauen von b in f als Vertrauenseinführer, mit dem die Transitivität von Vertrauen eingeschränkt werden kann.
- Falls b nicht der Ersteller des Dokuments ist: Das Vertrauen des Knoten x, von dem b das Dokument erhalten hat, in b: $S_x(T_x^D(x, b))$
- Für jeden Knoten $u \in V_b^D$, für den $T_b^D(a, u) \geq \text{Conf}(Z)$ (einschließlich f) :
  - $E_u(S_b(T_b^D(b, u)))$
    Dadurch, dass die Vertrauensbeziehungen von b mitgeschickt werden, können andere Knoten Vertrauenseinstufungen für a berechnen. Sie sind jeweils mit dem öffentlichen Schlüssel des vertrauten Knotens verschlüsselt. Im konkreten Beispiel bedeutet das, dass f nicht erfährt, welchen anderen Knoten b vertraut – dies ist nur unter Mitwirkung des jeweiligen Knotens möglich. f kann selbst nur einen einzigen Eintrag des Vektors von Vertrauensbeziehungen selbst entschlüsseln. Durch Anfügen aller Vertrauensbeziehungen von A entsteht ein erheblich erhöhter Übertragungsaufwand. Wird für die Verschlüsselung beispielsweise der RSA-Algorithmus verwendet, so entspricht die Datenmenge, die für einen Eintrag benötigt wird, der Länge des in diesem Algorithmus verwendeten Modulus[20] – typischerweise 1024 oder 2048 bit (128 bzw. 256 byte). Dieser Overhead ist jedoch nötig, um den Verzicht auf eine zentrale Speicherung von Vertrauensbeziehungen zu ermöglichen. Der Vergleich mit anderen Systemen, in denen Vertrauenswerte gespeichert werden (wie dem PGP-Web-of-Trust[21]), führt allerdings zu der Vermutung, dass die Anzahl an direkten Vertrauensbeziehungen sich im Rahmen hält und typischerweise nicht über den unteren zweistelligen Bereich hinausgeht. Die resultierende Nachrichtengröße bleibt also handhabbar. Der sich beim Empfänger ergebende Rechenaufwand kann aus dem gleichen Grund vernachlässigt werden; denn er ergibt sich im wesentlichen aus dem Nachvollziehen der Berechtigungen bei der Weiterleitung der Nachricht über den gewählten Pfad.

[20] Unter der Annahme, dass $S_b(T_b^D(b, u))$ eine geringere als die genannte Datenmenge beansprucht – dieser Wert ist jedoch durchaus erreichbar.

[21] Eine Analyse [Cede07] des PGP-Web-of-Trust hat ergeben, dass diejenigen PGP-Schlüssel, die zur größten Menge von Schlüsseln gehören, zwischen denen jeweils ein Vertrauenspfad besteht, im Durchschnitt 10,04 Signaturen tragen.

Zusätzlich wird die gesamte Nachricht M mit dem öffentlichen Schlüssel von f verschlüsselt. Gesendet wird also $E_f(M)$.

Ist das Vertrauen des Autors in den Zielknoten aus Sicht des weiterleitenden Knoten hingegen kleiner als die Vertraulichkeitseinstufung des Dokuments – es gilt also $T_b^D(a,f)$ <Conf(Z) –, so ist die Weiterleitung des Dokuments zunächst nicht möglich. Dies entspricht in Abbildung 4.3 dem Zustand „Weiterleitung nicht berechtigt“. In diesem Fall muss b zunächst feststellen, ob die verschlüsselte Vertrauensinformation, die er empfangen hat, aber nicht lesen kann, eine andere Einstufung rechtfertigt:
b schickt an f für jeden Vorgänger y von b im Pfad von a zu b den Vektor mit den Einträgen $E_u(S_y(T_y^D(y,u)))$ mit $u \in V_y^D$, wobei für u gilt $T_y^D(a,u)$ ≥Conf(Z) – das Dokument selbst jedoch noch nicht.

Kann f einen der Einträge mit seinem privaten Schlüssel entschlüsseln (ist also eine Vertrauensbewertung für f enthalten), entschlüsselt f diesen Eintrag $E_f(S_y(T_y^D(y,f)))$ und schickt ihn – verschlüsselt mit dem öffentlichen Schlüssel von b – an b.[22] Somit ist der Grad an Vertrauen von y in f bewiesen. Kann f den Eintrag nicht entschlüsseln, sind keine weiteren Aktionen möglich.
Erhält b eine Antwort, so gelangt er in den Zustand „Entschlüsselten Vertrauenswert empfangen“. b speichert den Wert $T_y^D(y,f)$ ab und wiederholt die Berechnung von $T_b^D(a,f)$. Der neue Wert ist in Abbildung 4.3 als T' bezeichnet. Ist dieser Wert nun größer als Conf(Z), kann b weiter vorgehen, als hätte er die Vertrauensbeziehung zwischen a und f selbst herstellen können. Ansonsten kann das Dokument endgültig nicht weitergeleitet werden.

4.4.2.6 Erhalt einer Nachricht

Erhält ein Knoten f eine Nachricht von seinem Vorgänger b, so entschlüsselt er diese. Sodann

- prüft er die Signatur von Dokument und Vertraulichkeitseinstufung $S_a(Z$,Conf$(Z))$.
- prüft er die Signatur aller Einstufungen als Vertrauenseinführer $S_x(I^D(x,y))$, die jeder Knoten x entlang des Pfades für seinen jeweiligen Nachfolger Y der Nachricht hinzugefügt hat, und speichert den Wert (einschließlich der Signatur) lokal ab.
- prüft er die Signaturen der Vertrauenswerte $S_x(T_x^D(x,y))$, die jeder Knoten x entlang des Pfades für seinen jeweiligen Nachfolger y der Nachricht hinzugefügt hat, und speichert diese lokal ab.
- entschlüsselt er den Wert $E_f(S_b(T_b^D(b,f)))$ und speichert den Vertrauenswert einschließlich der Signatur lokal ab.

[22] Möchte f seine Vertrauensbeziehung zu y nicht offenlegen, kann dieser Knoten auch auf das Senden einer Antwort verzichten.

Mit Hilfe der nun lokal gespeicherten Vertrauensinformationen berechnet f das Vertrauen des Dokumenterstellers A in f aus Sicht seines Vorgängerknotens b ($T_b^D(a, f)$) und erfährt so, ob die Weiterleitung des Dokuments durch b protokollkonform war.

4.4.2.7 Änderung von Vertrauenseinstufungen

Ändert sich das Vertrauen eines Knotens in einen anderen – oder ist eine festgelegte Zeitspanne seit der Erstellung vergangen – müssen noch zwischengespeicherte Vertrauenswerte überschrieben werden. Wird nach einer lokalen Änderung erstmals ein Dokument versandt, wird das vorgeschlagene Protokoll ohne Änderung durchgeführt. Allerdings wird im von A versendeten Vektor $E_U(S_a(T_a^D(a, u)))$ der entsprechende Eintrag mit einer Markierung versehen, die auf die Aktualisierung hinweist. Auch der neue Vertrauenswert wird jedoch nicht offengelegt. Stattdessen werden alle von a vorgenommenen Vertrauensbewertungen aus dem Speicher der Empfängerknoten gelöscht. Es ist denkbar, dass ein Knoten sich nicht an diese Vorgabe hält: Empfängt ein Knoten f den Vektor $E_u(S_a(T_a^D(A, U)))$ und soll den Vertrauenswert $T_a^D(a, u)$ entschlüsseln, könnte er zwar noch mit dem alten Wert antworten; der in der Signatur enthaltene Zeitstempel ermöglicht aber eine Überprüfung der Aktualität der Vertrauensbewertung.

Mit dieser Prozedur kann allerdings nicht immer verhindert werden, dass ein nicht mehr als ausreichend vertrauenswürdig betrachtetet Knoten ein Dokument erhält, da das Propagieren der neuen Vertrauensinformation durch das Netz eine gewisse Zeit in Anspruch nimmt und in der Zwischenzeit das Dokument weiter an nicht mehr vertrauenswürdige Knoten verbreitet werden kann.

4.5 Erfüllung der Anforderungen

Dieser Abschnitt analysiert, in welchem Maß die Anforderungen aus Abschnitt 4.4.2.1 durch das beschriebene Verfahren erfüllt werden können.

1. *Das Verfahren soll ohne zentrale Komponenten auskommen.*
   Beim Entwurf wurden keine zentralen Komponenten verwendet; es ist lediglich denkbar, dass die PKI, die als Grundlage des Systems nötig ist, zentrale Komponenten benötigt.

2. *Ein Dokument darf nur dann an einen Knoten b weitergeleitet werden, wenn (aus der Sicht des weiterleitenden Knotens a) das Vertrauen des Autors c des Dokuments in b (bezogen auf die Domäne des Dokuments Z) größer oder gleich der Vertraulichkeitseinstufung des Dokuments ist:* $T_a^D(c, b) \geq Conf(Z)$.
   Verhält der weiterleitende Knoten sich protokollkonform, ist die Erfüllung dieser Anforderung unproblematisch:

- Der weiterleitende Knoten kennt $\mathrm{Conf}(Z)$, denn dieser Wert wird dem Dokument angefügt und signiert. Durch das Vorhandensein einer PKI ist es dem weiterleitenden Knoten möglich, diese Signatur auch zu prüfen.
- Der weiterleitende Knoten a, der ein von b erstelltes Dokument an c weiterleiten will, muss auch $T_a^D(b, c)$ berechnen können. Dazu benötigt er
  - falls ein ausreichender Vertrauenspfad von b zu c über a selbst läuft: Das Vertrauen von b in a als Vertrauenseinführer. Entweder hat a diesen Vertrauenswert direkt von b erhalten – dann ist er von b signiert, seine Authentizität kann geprüft werden, und a muss lediglich noch sein eigenes Vertrauen in c kennen, um anhand des Vertrauensmodells $T_a^D(b, c)$ errechnen zu können. Oder es existiert ein Pfad von Vertrauenseinführerbewertungen[23] von b zu A. Diesen Pfad kennt A, denn jeder Knoten entlang des Pfades, über den das Dokument gesendet wurde, hängt an die Nachricht die Vertrauenseinführerbewertung für seinen Folgeknoten an und signiert diese. So muss a lediglich die Signaturen prüfen und kennt wiederum seine Bewertung als Vertrauenseinführer.
  - falls ein ausreichender Vertrauenspfad aufgrund der zwischengespeicherten Vertrauensbeziehungen zwischen anderen Knoten besteht: Die Bewertungen als Vertrauenseinführer entlang dieses Pfades bis zum letzten Knoten, der im Pfad vor c liegt, sowie dessen Vertrauen in C. Sofern diese Werte zwischengespeichert sind, hat b auch sie zusammen mit einer Signatur erhalten.
  - falls eine Abfrage höherer Vertrauenswerte bei c notwendig ist: Das Vertrauen von a in c. c sendet diese Einstufung an b; sie ist von a signiert, und b kann diese Signatur prüfen.
- Ist die Herstellung eines Vertrauenspfads so nicht möglich, wird das Dokument auch nicht weitergeleitet.

Je nach Anwendungsszenario können Dokumente – insbesondere wenn sie für die Verwendung durch den menschlichen Benutzer bestimmt sind – unabsichtlich offengelegt werden. Hier kann aus technischer Sicht lediglich eine Warnung erfolgen.

Ist ein weiterleitender, als vertrauenswürdig eingestufter Knoten ein Angreifer, kann der Empfänger die Protokollverletzung – also die Weiterleitung an einen Knoten, die nach dem Protokoll nicht zulässig war – erkennen: Er überprüft die Signatur der Vertraulichkeitseinstufung des Dokuments und berechnet – ebenso wie der Knoten, von dem er das Dokument erhalten hat – seine eigene Vertrauenseinstufung. Entfernt der Angreifer allerdings die Information über den Ersteller des Dokuments und gibt sich selbst als solcher aus, kann die protokollwidrige Weiterleitung nicht mehr entdeckt

[23] Dieser Pfad kann sich von einem Vertrauenspfad aus direkten Vertrauensbewertungen unterscheiden.

werden. Es handelt sich um ein prinzipielles Problem, denn mit Mitteln der Kryptographie kann zwar die Authentizität eines Dokuments nachgewiesen werden, das unerlaubte Entfernen der Authentifizierungsdaten lässt sich jedoch nicht verhindern. Eine Angriffserkennung wäre nur noch dadurch möglich, dass zwei inhaltlich identische Dokumente mit verschiedenen Autoren im Netz vorhanden sind – auch könnten beispielsweise digitale Wasserzeichen eingesetzt werden, um den tatsächlichen Autor rekonstruieren zu können. Beides ist jedoch nicht Gegenstand dieser Arbeit.

Für den Anwendungsfall der Empfehlungssysteme ist dieser Angriff jedoch unproblematisch: Enthält ein weitergeleitetes Dokument eine Bewertung für ein bestimmtes Objekt, so besteht die vertrauliche Information darin, dass der Autor des Dokuments dieses Objekt bewertet bzw. in seinem Besitz hat – ohne diesen Personenbezug ist die Bewertung eines Objekts grundsätzlich unkritisch. Entfernt ein Angreifer nun die Information über den Autor, so besteht auch kein Schutzbedarf mehr. Er könnte sich zwar selbst als Autor ausgeben; aus Sicht des Datenschutzes ist dies jedoch unproblematisch.

Die Preisgabe eines vertraulichen Dokuments kann im Übrigen nicht verhindert werden, wenn nur ein einzelner Knoten sich im Nachhinein als nicht vertrauenswürdig herausstellt.

3. *Aus Gründen des Datenschutzes sollten Vertrauensbeziehungen in einen Knoten nur mit Mitwirkung dieses Knotens offengelegt werden können.*
   Vertrauenseinstufungen werden mit dem öffentlichen Schlüssel des Knotens verschlüsselt, dem vertraut wird. Das asymmetrische Kryptographieverfahren, das eingesetzt wird, garantiert, dass nur mit dessem privaten Schlüssel eine Entschlüsselung der Vertrauensbeziehung möglich ist.

4. *Ein Knoten m kann $T_m^D(b,c)$ nicht unentdeckt höher als von b erlaubt wählen.*
   Diese Anforderung lässt sich zurückführen auf Anforderung 2. Die Informationen, die M zur Verfügung stehen, um zu beurteilen, ob $T_m^D(b,c) \geq \mathrm{Conf}(Z)$, stehen auch dem Nachfolgeknoten c zur Verfügung, so dass dieser überprüfen kann, ob die Weiterleitung gerechtfertigt war.

5. *Die Erlangung mehrerer Identitäten gibt einem Angreifer keinen Vorteil.*
   Solange kein Teilnehmer außerhalb der Gruppe von Identitäten, die durch den Angreifer erzeugt wurde, einer Identität innerhalb dieser Gruppe vertraut, werden keine Dokumente an diese Gruppe gesendet. Die Erzeugung multipler Identitäten ermöglicht die Erzeugung von längeren Vertrauenspfaden einerseits und von mehr Vertrauenspfaden andererseits. Längere Vertrauenspfade führen lediglich zu reduziertem Vertrauen; zusätzliche Vertrauenspfade bleiben wirkungslos, da nur derjenige

Pfad berücksichtigt wird, der zur höchsten Vertrauenseinstufung führt. Insbesondere kann ein Angreifer sich nicht am Beginn eines Vertrauenspfads positionieren – an dieser Stelle hat noch keine multiplikative Verkleinerung des Vertrauenswerts stattgefunden –, denn dies würde Vertrauen durch den Autor selbst erfordern. Während das Protokoll für die Anwendung transitiven Vertrauens sicherstellt, dass die Verwendung mehrerer Identitäten keinen Vorteil bietet, ist dies für die erforderlichen direkten Vertrauensbeziehungen durch Annahme A3 erfasst.

6. *Vertrauenseinstufungen können jederzeit geändert werden.*
   Das Verfahren für die Veränderung von Vertrauensbeziehungen ist in Abschnitt 4.4.2.7 beschrieben.

7. *Ständige Erreichbarkeit der Netzteilnehmer sollte nicht erforderlich sein.*
   Das Verfahren erfordert keine länger anhaltenden Kommunikationsbeziehungen, als sie für den Dokumentenaustausch selbst nötig sind: Die benötigte zusätzliche Kommunikation findet unmittelbar vor dem Austausch des jeweiligen Dokuments statt.

8. *Das Verfahren soll auch skalierbar sein; der Kommunikationsaufwand soll im Vergleich zu einem Peer-to-Peer-System ohne den Schutz der Vertraulichkeit höchstens um einen konstanten Faktor steigen.*
   Im beschriebenen Verfahren werden bei der Weiterleitung eines Dokuments keine zusätzlichen Nachrichten ausgetauscht; die für das Verfahren benötigten Informationen werden Nachrichten angehängt, die ohnehin ausgetauscht werden. Einzige Ausnahme bildet die Abfrage nach höheren Vertrauenswerten, für die jeweils höchstens zwei zusätzliche Nachrichten gebraucht werden (Anfrage und ggf. Antwort). Der Umfang der übertragenen Nachrichten wird deutlich erhöht: Pro durchlaufenem Knoten werden die Vertrauenseinstufungen dieses Knotens hinzugefügt. Jedoch ist die Anzahl dieser Vertrauenseinstufungen nicht von der Anzahl der Knoten, die das Verfahren nutzen, abhängig; stattdessen ist nur von einer geringen Anzahl solcher Vertrauensbeziehungen pro Knoten auszugehen[24]. Die Nachrichtengröße wird insbesondere nicht multiplikativ erhöht, sondern es wird ein von Nachrichten- und Netzgröße unabhängiger Overhead hinzugefügt. Die grundsätzliche Umsetzbarkeit des Verfahrens wurde durch eine prototypische Implementierung auf Basis des SESAM-Systems [CDHS+05] im Rahmen einer Studienarbeit [Wund07] gezeigt.

4.5.1 Fazit

Das Konzept des Vertrauens wird nicht nur in der Forschung diskutiert, sondern ist auch von erheblicher praktischer Relevanz. Im vorliegenden Kapitel wurden die Grundlagen des

[24] Im PGP-Web-of-Trust liegt diese Anzahl ungefähr bei 10, vgl. Fußnote 21

Phänomens Vertrauen aus der Sicht verschiedener Wissenschaften aufgezeigt und sodann Mechanismen zur Vertrauenserzeugung für Anwendungen in Rechnernetzen betrachtet. Der Schwerpunkt wurde dabei auf Reputationssysteme gelegt, die in der Folge auch aus juristischer Sicht betrachtet wurden. Es zeigt sich, dass die Rechtmäßigkeit von Bewertungen, die im Rahmen eines solchen Systems abgegeben werden, in vielen Fällen der Abwägung von grundrechtlich geschützten Positionen des Bewertenden und des Bewerteten bedarf. Die Erkenntnisse, die für Reputationssysteme gewonnen wurden, lassen sich überwiegend auch auf Empfehlungssysteme übertragen. Schließlich wurde ein Verfahren vorgestellt, mit dem – ausgehend von Vertrauensbeziehungen zwischen Nutzern – die Weitergabe von Dokumenten in einem Peer-to-Peer-Netz auf vertrauenswürdige Knoten beschränken lässt.

Die Ergebnisse dieses Kapitels sind in Teilen in [SoDZ06] und [SoZi06] veröffentlicht.

Kapitel 5

Entwurf eines Empfehlungssystems

Das vorliegende Kapitel befasst sich mit der konkreten Umsetzung eines Empfehlungssystems.

Das Kapitel ist wie folgt gegliedert: Zunächst werden die Inhalte von Bewertungsdokumenten und Nutzerprofilen betrachtet, die in allen verfolgten Ansätzen verwendet werden. Anschließend werden Aufgaben und Anforderungen der entworfenen Systeme identifiziert sowie eine Architektur präsentiert, in die sich die dargestellten Verfahren einordnen lassen. Den Abschluss des Kapitels bildet die Darstellung einzelner Komponenten dieser Architektur: Overlay-Netz und Datenspeicherung als gemeinsame Komponenten sowie die darauf aufbauenden Anwendungen: Nutzerbasiertes Kollaboratives Filtern, die verteilte Speicherung von Bewertungsdokumenten sowie objektbasiertes Kollaboratives Filtern.

5.1 Bewertungsdokumente

Die Bewertung eines Objekts durch einen Nutzer wird in einem *Bewertungsdokument* zusammengefasst. Das Format der verwendeten Bewertungsdokumente wurde bislang in der Literatur nicht diskutiert. Für die vorliegende Arbeit sollen Bewertungsdokumente mit folgendem Inhalt verwendet werden:

- Identifikator des Bewertungsdokuments
- Identifikator des bewerteten Objekts
- Erstellungszeitpunkt des Bewertungsdokuments
- Bewertung in Form einer Bewertungszahl a zwischen 0 und 1
- Identität (Pseudonym) des Dokumentautors
- Einordnung des Objekts in eine Kategorie[1] (optional, beliebig oft)

[1] Kategorien können mit den im letzten Abschnitt beschriebenen Domänen identisch sein; die Unterteilung in Kategorien ist aber potentiell von höherer Granularität.

- Klartextbezeichnung des Objekts (optional)
- Identifikatoren verwandter Objekte (optional)
- Weitere Identifikatoren des gleichen Objekts (optional)
- Textuelle Bewertung und Beschreibung des Objekts (optional)
- Signatur (optional)

Durch den Empfehlungsalgorithmus wird im Rahmen des Kollaborativen Filterns dabei nur die Bewertungszahl und die Identität des Dokumentautors berücksichtigt. Daneben wird von der Anwendung die Signatur des Dokuments geprüft. Die weiteren Werte können für die Darstellung in der Nutzerschnittstelle verwendet werden.

Die Bewertungsdokumente eines Nutzers bilden zusammengenommen sein *Nutzerprofil*; für jedes Objekt, das ein Nutzer bewertet hat, enthält es genau ein Bewertungsdokument.

5.2 Aufgaben, Anwendungsfälle und Anforderungen

Als Anforderungen an ein Empfehlungssystem ergeben sich, dass jeder Benutzer

- unter einer bestimmten ID eines Objekts ein Bewertungsdokument in das Empfehlungssystem einspeisen können soll – jedoch nicht mehr als eines. Hintergrund dieser Anforderung ist, dass die Gesamteinschätzung eines Objekts durch einen Nutzer als Grundlage der Empfehlungserzeugung dient und nicht etwa die Einschätzung nach jeder einzelnen Benutzung des Objekts.
- dieses Bewertungsdokument zu einem späteren Zeitpunkt ändern oder löschen können soll. Ein bestehendes Bewertungsdokument kann also durch ein neues ersetzt werden.
- zur gegebenen ID eines Objektes alle zu diesem Zeitpunkt im Netz verfügbaren Bewertungsdokumente abrufen können soll, die sich auf diese ID beziehen.
- auf Grundlage seiner bisherigen Objektnutzungen und -bewertungen Empfehlungen für weitere Objekte erhalten können soll. Dies umfasst zum einen das *Auffinden* von Objekten, die für einen Nutzer von Interesse sein könnten, zum anderen die *Vorhersage* der Bewertung, die ein Nutzer einem Objekt voraussichtlich geben wird.

Darüber hinaus soll das System skalierbar sein, der benötigte Speicher- und Kommunikationsaufwand also auch für große Teilnehmer- und Bewertungszahlen beherrschbar bleiben.

5.2.1 Anforderungen an Sicherheit und Datenschutz

Neben die funktionalen Anforderungen treten weitere Anforderungen, die Sicherheit und Datenschutz betreffen. Sicherheitsanforderungen betreffen die Sicherheit vor Angriffen, die Einfluss auf die Funktionalität des Systems nehmen. Datenschutzanforderungen betreffen die Sicherheit vor Angriffen, die auf die Aufdeckung personenbezogener Daten von Nutzern des Systems zielen. Für die Sicherheit des Systems ergeben sich als Anforderungen konkret:

- Die Verfügbarkeit des Systems ist zu gewährleisten; Denial-of-Service-Angriffe (DoS-Angriffe) sollen erschwert werden.
- Das Zurückziehen einzelner Bewertungsdokumente – außer durch den Autor – ist zu verhindern.
- Die Integrität und Authentizität der gespeicherten Bewertungsdokumente ist zu gewährleisten.
- Das massenweise Einspeisen von Bewertungsdokumenten soll erschwert werden.
- Der Einsatz eines Reputationssystems soll möglich sein. Diese Anforderung betrifft indirekt die Sicherheit des Systems, denn sie ermöglicht die Sanktionierung unerwünschten Verhaltens durch Verlust von Reputation.

Aus Sicht des Datenschutzes ist weiterhin die Identität beteiligter Teilnehmer zu schützen. Beim Abruf von Bewertungsdokumenten soll die Identität des abrufenden Teilnehmers überhaupt nicht offengelegt werden; beim Einfügen solcher Dokumente in das System ist die Identität des Einfügenden zu schützen, soweit dies unter Wahrung der Anforderungen an Funktionalität und Sicherheit möglich ist. Der Verzicht auf die Erhebung personenbezogener Daten bedeutet auch, dass die Sperrung eines Nutzers beispielsweise bei missbräuchlicher Verwendung des Systems erschwert wird.

Da – wie sich in diesem Kapitel zeigen wird – die Erfüllung von Datenschutzanforderungen in einem erheblich erhöhten Aufwand sowie reduzierter Sicherheit resultiert, sollen die entsprechenden Maßnahmen optional sein.

5.3 Gliederung und Architektur

Abbildung 5.1 zeigt die Architektur, die dem Entwurf, wie er in diesem Kapitel beschrieben wird, zugrunde liegt.

Alle beschriebenen Ansätze bauen auf einem Netz auf, in dem die Protokollfamilie TCP/IP zum Einsatz kommt – dies ermöglicht auch den problemlosen Einsatz im Internet. Auf diese Protokollfamilie soll an dieser Stelle nicht eingegangen werden; relevant ist für die

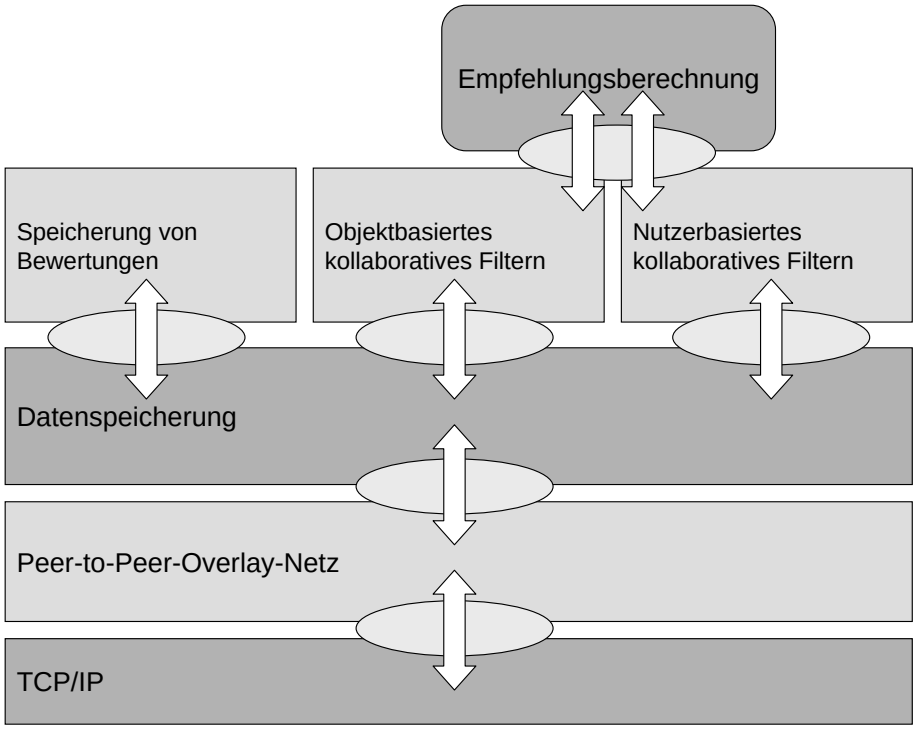

Abbildung 5.1: Überblick der entworfenen Architektur

Zwecke der vorliegenden Arbeit lediglich, dass die Kommunikation zwischen Anwendungen ermöglicht wird, die auf mit dem Internet verbundenen Endgeräten ausgeführt werden.

Darauf aufgebaut wird ein Peer-to-Peer-Overlay-Netz. Die Anwendungen „Objektbasiertes Kollaboratives Filtern“ und „Speicherung von Bewertungen“ erfordern ein strukturiertes Overlay-Netz nebst darauf aufbauender Datenspeicherungsschicht. Dieses und die darauf aufbauende Datenspeicherung werden am Beispiel Chord in Abschnitt 5.4 beschrieben. Die betrachteten strukturierten Overlay-Netze erlauben das Senden einer Nachricht an einen Knoten, der für einen bestimmten Schlüssel zuständig ist. Die Anwendung „Nutzerbasiertes Kollaboratives Filtern“ kann grundsätzlich auch auf einem unstrukturierten Overlay-Netz und einer zugehörigen Datenspeicherungsschicht aufbauen. Tatsächlich umgesetzt wird auch diese Anwendung aber auf Basis eines strukturierten Netzes.

Die Datenspeicherungsschicht (Abschnitt 5.5) definiert die verteilte Speicherung von Dokumenten aufbauend auf einem Overlay-Netz. Sie bietet das Einfügen von Dokumenten unter einem bestimmten Schlüssel, die Suche nach Dokumenten unter Verwendung dieses Schlüssels sowie die Löschung derselben an.

Auf der Datenspeicherungsschicht baut einerseits die verteilte Speicherung von Bewertungsdokumenten auf (Abschnitt 5.7), andererseits zwei Anwendungen, die Kollaboratives Filtern ermöglichen. Zum einen handelt es sich um einen nutzerbasierten Ansatz, der in Abschnitt 5.6 beschrieben und in Abschnitt 5.6.6 zu einem kombinierten Ansatz erweitert

wird. Andererseits wird ein objektbasierter Ansatz zur Empfehlungserzeugung entwickelt (Abschnitt 5.8). Die beiden letztgenannten Anwendungen bieten die gleichen Schnittstellen: Sie erlauben dem Nutzer, eine Bewertung abzugeben, sich eine Bewertung für ein bestimmtes Objekt vorhersagen zu lassen oder eine Liste relevanter Objekte zu generieren. Die Anwendungen greifen dabei auf einen Empfehlungsalgorithmus zurück, der Empfehlungen auf das Basis einer Menge von Bewertungen errechnet. Die Suche und Anzeige von Bewertungsdokumenten wird durch diese beiden Anwendungen nicht realisiert, sondern erfordert eine Kombination mit der Anwendung „verteilte Speicherung von Bewertungsdokumenten“.

5.4 Overlay-Netz

In dieser Schicht wird der effiziente Zugriff auf bestimmte Datenblöcke (insbesondere können dies Dokumente sein) ermöglicht, die einem bekannten Identifikator (Schlüssel) zugeordnet sind. Für diesen Zweck sind in der Literatur diverse Peer-to-Peer-Overlay-Netze vorgeschlagen worden. Auch, wenn typische Implementierungen oft gleichzeitig weitere Teilaufgaben lösen, soll an dieser Stelle lediglich das effiziente Routing von Anfragen zu einem Kommunikationspartner betrachtet werden.

Als *unstrukturiertes* Netz kann ein beliebiges Netz zum Einsatz kommen, das Nachbarschaftsbeziehungen zu anderen Knoten finden und aufrecht erhalten kann; zu nennen ist beispielsweise Gnutella [Ripe01].

Da *strukturierte* Netze diese Eigenschaft auch erfüllen, dabei aber zusätzliche Vorteile – insbesondere die Möglichkeit des effizienten Auffindens von Ressourcen – bieten (vgl. Abschnitt 2.1.5.2), liegt der Schwerpunkt jedoch auf der Anwendung solcher Netze. Konkret wird die DHT (Distributed Hashtable) Chord [SMKK$^+$01] mit rekursiven Lookup-Operationen verwendet. Soll eine Nachricht an den für einen Schlüssel zuständigen Knoten gesendet werden, so sendet der Senderknoten diese ab, und sie wird ohne seine weitere Mitwirkung bis zum zuständigen Empfänger geleitet.

Der Grund für die Auswahl von Chord liegt zum einen am geringen Such- sowie Speicheraufwand, der jeweils in $O(\log N)$ (N Anzahl der Knoten) liegt. Zum anderen ist Chord ein sehr gut untersuchtes System, und das Vorhandensein verlässlicher und getesteter Implementierungen erlaubt eine zuverlässige Evaluierung.

Während auf der Overlay-Ebene für das Routing vollständig auf die ursprüngliche Version des Chord-Protokolls zurückgegriffen werden kann, ist die Vergabe der Identitäten eine besondere Herausforderung. Bedingt wird dies durch die Existenz des Sybil-Angriffs (vgl. Abschnitt 2.1.7). Dennoch soll die Identitätsvergabe – als Aufgabe der Overlay-Schicht – nicht betrachtet werden; es werden lediglich die folgenden Annahmen getroffen:

- Die Identitäten werden in der Weise zugewiesen, dass

  - sie langfristig gültig sind,
  - ein Knoten seine Identität nicht gezielt wählen kann und
  - der Sybil-Angriff vermieden wird.

- Ein Knoten kann seine Identität beweisen, die ihm auf Overlay-Ebene zugewiesen ist.

Lösungsansätze, die dies gewährleisten sollen, sind in der Literatur beschrieben und in Abschnitt 2.1.7 zusammengefasst.

Zusätzlich wird eine Public-Key-Infrastruktur (PKI) vorausgesetzt. Public-Key-Infrastrukturen für Peer-to-Peer-Netze wurden in der Literatur [WöWü05] vorgeschlagen; auch kann das PGP Web of Trust für diesen Zweck eingesetzt werden.[2]

5.5 Datenspeicherung

Oberhalb der Overlay-Ebene wird die verteilte Datenspeicherung betrachtet. Der Schwerpunkt liegt dabei in der Verwendung einer DHT (Distributed Hashtable) als Overlay-Schicht (vgl. dazu Abschnitt 2.1.5.2). Zunächst wird die grundlegende Funktionsweise vorgestellt; Erweiterungen, die für einzelne Ansätze der folgenden Abschnitte nötig sind, werden dort erörtert.

Der folgende Entwurf beschreibt die Verwendung einer Datenspeicherungsschicht, die auf einer DHT basiert.[3]

Die Schnittstelle dieser Datenspeicherungsschicht umfasst lediglich die Operationen `void`[4] *einfügen(Schlüssel, Dokument)*, `void` *löschen(Schlüssel, Dokument-Identifikator, Signatur)*, `Dokument` *suchen(Schlüssel)* und `Dokument` *suchen(Schlüssel, Sucheinschränkung)*. Die letztgenannte Operation erlaubt es, bei Speicherung mehrerer Dokumente unter einem Schlüssel eine Einschränkung der gesuchten Dokumente vorzunehmen.

[2] Kann keine bestehende PKI eingesetzt werden, so kann stattdessen folgender, einfacher Mechanismus verwendet werden:

- Jeder Knoten generiert lokal ein Schlüsselpaar aus öffentlichem und privatem Schlüssel.
- Den öffentlichen Schlüssel speichert er (unter Verwendung der im folgenden Abschnitt beschriebenen Datenspeicherungs-Schicht) auf den für seine Identität zuständigen Knoten. Die speichernden Knoten lassen sich die zugewiesene Identität beweisen. Jeder dieser Knoten signiert daraufhin den öffentlichen Schlüssel und sendet ihn an den Schlüsselinhaber zurück.
- Ein prüfender Knoten betrachtet die Signatur als gültig, sobald er mindestens einem der Unterzeichner des öffentlichen Schlüssels vertraut; anderenfalls führt er selbst eine Überprüfung der Identität des Unterzeichners durch.

[3] Für unstrukturierte Overlay-Netze kann stattdessen eine einfache Datenspeicherungsschicht eingesetzt werden. Jeder Knoten ist für die Speicherung der eigenen Dokumente verantwortlich. Zusätzlich könnten diese bei k Nachbarknoten gespeichert werden, die über jede Aktualisierung informiert werden und die Dokumente weiterhin bereithalten, wenn deren Ersteller das Netz verlässt. Diese Form der Datenspeicherung ist somit zwar möglich, soll aber nicht Gegenstand der vorliegenden Betrachtung sein.

[4] Stattdessen ist auch die Rückgabe eines boolschen Werts möglich, der in Abhängigkeit vom Erfolg der Operation gesetzt wird.

Als Schlüssel werden dabei in allen Fällen die in Abschnitt 2.2.2 diskutierten Identifikatoren für Objekte herangezogen. Dabei können zwar grundsätzlich Probleme entstehen – der Hersteller könnte beispielsweise die ID des Objekts und somit die Auswahl der zuständigen Knoten beeinflussen. Die konkrete Lösung des Problems hängt von der Anwendungsdomäne und den konkret verwendeten Identifikatoren ab. Bei Schemata wie ISBN kann der Hersteller (Verlag) den Identifikator nicht frei wählen, wodurch diese Problematik bereits gelöst wird.

Durch Anwendung einer kryptographischen Hash-Funktion wird aus dem Identifikator – wie in [SMKK+01] vorgesehen – jeweils die Chord-ID des Inhalts gewonnen. Für einen Schlüssel zuständig ist dann der Nachfolgerknoten der Chord-ID des Inhalts *Nachfolger(Chord-ID(Schlüssel)) = Nachfolger(hash(Schlüssel))*. Ein eingefügtes Dokument wird mit Hilfe des Chord-Routing zu demjenigen Knoten geroutet, der für den angegebenen Schlüssel zuständig ist.

Wie beim DHASH-Verfahren (vgl. Abschnitt 2.1.5.2) wird ein eingefügtes Dokument nicht nur auf diesem zuständigen Knoten K, sondern auch auf $r-1$ weiteren Nachfolgerknoten gespeichert; Nachfolgerknoten sind dabei die im Chord-Ring folgenden Knoten, nicht die $r-1$ ersten Einträge der finger table. Sobald der zuständige Knoten das Dokument erhalten hat, leitet er es also an seinen Nachfolgerknoten weiter. Zusätzlich zum übertragenen Dokument enthält die Nachricht einen Hop-Zähler, den K auf r setzt. Der Nachfolgerknoten von K leitet die Nachricht wiederum an seinen Nachfolger weiter und dekrementiert dabei den Hop-Zähler um 1. Die Nachricht wird nicht mehr weitergeleitet, sobald der Hop-Zähler den Wert 1 erreicht hat. Die so erzielte redundante Datenspeicherung soll als *lokale* Redundanz bezeichnet werden.

Betrachtet man auch Sicherheitsüberlegungen, so ist dieser Schutz jedoch nicht ausreichend, da der eigentlich zuständige Knoten auf Anfrage auch falsche Nachfolger liefern könnte (vgl. [SiMo02]). Gelöst werden kann dieses Problem, indem die redundante Speicherung zusätzlich an voneinander im Chord-Ring weit entfernten Knoten vorgenommen wird. Bereits der Ausgangsknoten sendet das Dokument im vorgeschlagenen System nicht nur an den Knoten *Nachfolger(hash(Schlüssel))*, sondern auch *Nachfolger(hash(i | Schlüssel) mod k)*, mit $i \in \{2, \ldots, s\}$, k Größe des Wertebereichs der Hashfunktion, $|$ Konkatenationsoperator. Für i ist dabei ein Feld fester Länge vorgesehen, so dass i nicht als Teil des eigentlichen Schlüssels aufgefasst werden kann. Auf diesem Weg kann s-fache Redundanz (die auch als *globale* Redundanz bezeichnet werden kann) erzielt werden. Der Kommunikations- und Speicheraufwand des Systems wird somit allerdings um den Faktor s erhöht. Bei Ausfall eines Knotens hat dessen nun neu zuständiger Nachfolger die Daten bereits gespeichert und muss lediglich die erneute Herstellung der lokalen (r-fachen) Redundanz sicherstellen. Lediglich, wenn ein Knoten überhaupt keine Antwort liefert – und sich so auch seine Nachfolger nicht ermitteln lassen – oder eine Überprüfung zurückgelieferter Ergebnisse er-

forderlich ist, wird tatsächlich auf den entfernten speichernden Knoten zurückgegriffen; ansonsten werden Anfragen nur an den ersten zuständigen Knoten gesandt.

Auf den im DHASH-Verfahren vorgeschlagenen Caching-Mechanismus soll verzichtet werden; im Vergleich zu klassischen Filesharing-Anwendungsszenarien ist bei Empfehlungssystemen mit einer häufigeren Aktualisierung der gespeicherten Daten zu rechnen: Eine Datei (beispielsweise ein Film oder ein Lied) kann sich zwar ändern; im Fall von bewerteten Objekten ist aber das Hinzufügen neuer Bewertungen sogar der Regelfall. Der Aufwand, den Cache aktuell zu halten, könnte so zu groß werden. Auch auf die in [DBKK+01] weiterhin vorgesehene Lastverteilung durch Aufteilung von Dokumenten bzw. Dateien kann für den Anwendungsfall des Empfehlungssystems zunächst verzichtet werden, da die meisten der abgerufenen Dokumente klein sind und die durch andere Verfahren erreichbare Verteilung der Systemlast somit ausreicht.

Zusammengefasst sendet ein Knoten, der ein Dokument einfügen will, dieses an die s zuständigen Knoten, die sich aus der Anwendung der Hash-Funktion auf den verwendeten Schlüssel ergeben. Jeder dieser Knoten sendet das Dokument an seine r Nachfolgerknoten. Hat einer der Empfänger bereits ein Bewertungsdokument für das entsprechende Objekt erhalten, dessen Autor identisch zum neu erhaltenen Dokument ist, so prüft er, ob dieses neu eintreffende Dokument neuer ist. Ist dies der Fall, wird das bestehende Dokument durch das neue ersetzt. Andernfalls wird die entsprechende Nachricht ignoriert.

Jeder Knoten, dem ein Vorgängerknoten zur Speicherung erhalten hat oder an einen Nachfolgerknoten zur Speicherung gesendet hat, prüft periodisch, ob dieser Knoten noch erreichbar ist; diese Prüfung wird ohnehin durch das Chord-Protokoll zur Verfügung gestellt. Ist der Vorgänger- bzw. Nachfolgerknoten nicht mehr erreichbar, sendet der jeweilige Knoten an den neuen Vorgänger bzw. Nachfolger das gespeicherte Dokument. Wie bereits beim DHASH-Verfahren werden einem neu beitretenden Knoten Dokumente zur Speicherung übermittelt, für die er zuständig wird. Da das Chord-Protokoll den Nachfolgerknoten des neu beitretenden Knotens über dessen Beitritt informiert, kann dieser prüfen, ob sein neuer Vorgänger für ein bei ihm gespeichertes Dokument zuständig wird. Ist dies der Fall, übermittelt er das entsprechende Dokument an den Vorgänger. Ist auf diesem Weg eine $r+1$-fach redundante Speicherung des Dokuments erreicht, wird der Knoten, der im Chord-Ring von dem für den Schlüssel eigentlich zuständigen Knoten am weitesten entfernt ist, benachrichtigt und löscht das Dokument lokal.

Ein Knoten, der ein Dokument sucht (Operation *suchenSchlüssel*), sendet eine Anfrage an den Knoten, der für hash(Schlüssel) zuständig ist. Erhält er nach einem Zeitraum Δt keine Antwort, so sendet er die Anfrage an den Knoten, der für hash(i | Schlüssel) zuständig ist. Dies wird für $i = 2, \ldots, s$ ausgeführt, sofern der Knoten vor Abschluss dieses Vorgangs keine Antwort erhält.

5.5.1 Löschung von Daten

Durch die Verwendung elektronischer Signaturen gestaltet die Löschung von Dokumenten sich einfach. Die Lösch-Operation erfordert den Schlüssel des zu löschenden Dokuments, um die speichernden Knoten auffinden zu können, sowie den Identifikator des entsprechenden Dokuments. Zusätzlich wird eine Signatur benötigt, die mit dem gleichen privaten Schlüssel erstellt wird wie die Signatur des zu löschenden Dokuments. Realisiert wird die Löschung eines Dokuments innerhalb der Datenspeicherungsschicht, indem ein Dokument erstellt wird, das einen Löschbefehl enthält und dann die Operation `void` einfügen(Schlüssel, Dokument) verwendet wird. Ein Knoten, der eine solche Löschanforderung erhält, prüft, ob er das zu löschende Dokument gespeichert hat; dann prüft er die Signatur der Löschanforderung und überprüft, ob sie mittels des gleichen privaten Schlüssels erstellt wurde wie die Signatur des Dokuments selbst. Ist dies der Fall, löscht er das Dokument.

Diese Art der Löschung ist unabhängig von den verwendeten Identitäten (z.B. Pseudonymen), da sie lediglich die Kenntnis des privaten Schlüssels erfordert.

5.5.2 Datenschutz in der Datenspeicherungsschicht

Aus Sicht des Datenschutzes gilt es, zwei Problemkreise zu unterscheiden: Einerseits soll der Nutzer beim Abrufen von Bewertungsdokumenten geschützt werden, andererseits soll auch die Profilbildung über Personen vermieden werden, die Bewertungsdokumente veröffentlichen.

5.5.2.1 Datenschutz beim Abruf von Dokumenten

Beim Abruf von Dokumenten ist im Wesentlichen problematisch, dass

- der für die Speicherung von Dokumenten mit einem bestimmten Schlüssel zuständige Knoten die Chord-IDs aller anfragenden Knoten erfährt, sofern zur Übertragung dieser Dokumente jeweils eine direkte Verbindung zu diesen anfragenden Dokumenten hergestellt wird. Beispielsweise handelt es sich bei diesen Dokumenten um alle Informationen zu einem bestimmten bewerteten Objekt oder das Nutzerprofil eines bestimmten Nutzers. Das Problem kann durch Caching reduziert, aber nicht gelöst werden.
- Suchanfragen eines bestimmten Knoten im ersten Schritt nur an eine begrenzte Anzahl anderer Knoten geleitet werden – diejenigen, die in der finger table dieses Knotens enthalten sind – und somit eine partielle Profilbildung ermöglicht wird: Jeder Knoten in der finger table kennt einen Teil der Suchanfragen des suchenden Knotens.

Die Lösung des ersten Problems liegt darin, keine direkte Verbindung auf Transportschicht zwischen dem speichernden und dem anfragenden Knoten herzustellen. Da im

Overlay-Netz rekursive Lookups (vgl. Abschnitt 2.1.5.2) verwendet werden, ist ein Weg, diese direkte Kommunikation zu vermeiden, bereits vorgezeichnet: Die Antwort – also das oder die gewünschten Dokumente – wird auf dem gleichen Pfad zurückübertragen wie die Anfrage selbst. Hierbei merkt sich jeder Knoten, der eine Anfrage weiterleitet, für eine beschränkte Zeitspanne den Knoten, von dem er diese erhalten hat, und sendet die später eintreffende Antwort an diesen zurück. Die Identität des Anfragenden braucht somit in der Anfrage nicht mehr enthalten zu sein. Der Kommunikationsaufwand steigt im Vergleich zu dem in [DBKK+01] vorgeschlagenen Caching-Verfahren höchstens linear an (die Anzahl versendeter Nachrichten entspricht dem doppelten der Pfadlänge von anfragendem zu antwortendem Knoten, die wiederum in $O(\log N)$ mit N Anzahl der teilnehmenden Knoten liegt). Verlässt jedoch einer der beteiligten Knoten zwischen Anfrage und Antwort das Peer-to-Peer-System, muss die Anfrage wiederholt werden. Es wird auch keine Sicherheit gegen einen Angreifer erzielt, der das gesamte System beobachten kann. Kooperieren mehrere Knoten auf dem Pfad einer Suchanfrage, besteht zudem das Risiko, dass der Pfad rekonstruiert werden kann.

Für das zweite Problem sind mehrere Lösungen denkbar:

- Das normale Chord-Routing könnte verwendet werden – jedoch wie bereits in der Lösung des ersten Problems, ohne den Anfragen Absenderinformationen beizufügen. Jeder Knoten entlang des Anfragepfads wüsste somit zwar, von wem er die Anfrage erhalten hat, nicht aber, wer der ursprüngliche Absender ist. Nachteil dieser Lösung ist jedoch die Möglichkeit, aufgrund der Funktionsweise des Chord-Routings Rückschlüsse darauf zu ziehen, ob ein Knoten eine Nachricht lediglich weitergeleitet hat oder er selbst der Absender ist. In [O'Va04] wird gezeigt, dass ein passiver Angreifer bei Verwendung des Chord-Protokolls zwar im Durchschnitt nur eine Menge von $\frac{n}{12}$ Knoten (mit n = Anzahl der Teilnehmer des Chord-Netzes) bestimmen kann, unter denen sich der Absender befindet. Diese Menge verkleinert sich jedoch, falls der Angreifer sich innerhalb des Chord-Rings weit vom zuständigen Knoten entfernt befindet oder wenn mehrere Knoten zusammenarbeiten. Andererseits wird die Menge durch Caching erheblich vergrößert.

- Auf dem ersten Hop könnten Anonymisierungsverfahren wie Crowds [ReRu98] oder Onion Routing [ReSG98] angewandt werden. Die Identität des Anfragenden kann mit diesen Verfahren zuverlässig geschützt werden, doch erhöht sich – in Abhängigkeit von der gewünschten Sicherheit – der Kommunikationsaufwand beträchtlich.

Zusammengefasst bestehen also beim Einfügen eines Bewertungsdokuments drei Möglichkeiten:

- Der suchende Knoten bedient sich des normalen Chord-Routings und fügt seine Absenderinformationen bei. Die Antwort wird direkt zum suchenden Knoten gesandt.

- Der suchende Knoten bedient sich des normalen Chord-Routings und fügt keine Absenderinformationen bei. Die Antwort wird auf dem gleichen Pfad zurückgesandt wie die Anfrage.

- Der suchende Knoten sendet seine Anfrage mittels des in [SyGR97] beschriebenen Verfahrens an den ersten Knoten auf dem Pfad zum für den gesuchten Schlüssel zuständigen Knoten. Die Antwort wird dann ebenfalls an diesen Knoten gesendet, der sie daraufhin ebenfalls mittels Onion Routing an den ursüprünglich anfragenden Knoten zurückliefert.

Die im Rahmen der vorliegenden Arbeit entstandene Implementierung greift auf die erste Variante zurück; es handelt sich um die effizienteste, die allerdings das geringste Datenschutzniveau gewährleistet. Es ist jedoch möglich, das Verfahren konfigurierbar zu gestalten, so dass für besonders sensible Anfragen eine der beiden anderen Lösungen verwendet werden kann. Das entsprechende Verfahren ist aus diesem Grund einfach austauschbar.

5.5.2.2 Datenschutz für Ersteller von Bewertungen

Für die Ersteller von Bewertungen stellt sich die Datenschutzproblematik anders dar: Bewertungsdokumente werden über einen längeren Zeitraum gespeichert. Auch, wenn keine Suche nach Autoren, sondern lediglich nach bewerteten Objekten möglich ist, so kann durch systematisches Abfragen von Bewertungsdokumenten, wenn auch unter sehr hohem Aufwand, doch ein Profil eines Autors erstellt werden: Ein Angreifer fragt dazu alle ihm bekannten Objekte ab und erhält zu jedem Objekt unter anderem eine Liste bewertender Autoren. Durch Neusortieren dieser Liste können Profile erstellt werden. Zudem kann auf diese Art und Weise nicht nur die Tatsache festgestellt werden, dass ein Nutzer sich für ein Objekt interessiert, sondern auch, wie er dieses bewertet hat.

Das Problem kann auf verschiedenen Ebenen des in Abschnitt 3.1.1.1 beschriebenen Modells gelöst werden:

- Auf der Identitätsebene. Dies bedeutet, dass die Identität des Bewertenden aus der Bewertung nicht erkennbar sein sollte. Während diese Anforderung grundsätzlich unproblematisch zu erfüllen ist, stellt sie im Zusammenhang mit der Verwendung von Reputationssystemen eine Herausforderung dar. Wie dies dennoch ermöglicht werden kann, wird in Abschnitt 5.7.1 erörtert.

- Auf der Empfängerebene. Hierzu könnte das in Abschnitt 4.4.2 beschriebene System verwendet werden, um Bewertungsdokumente nur vertrauenswürdigen Empfängern zugänglich zu machen. Darauf soll jedoch zunächst verzichtet werden, da die Profilbildung durch andere Maßnahmen hinreichend erschwert werden kann.

- Auf der Zweckebene. Die Verwendung eines Dokuments zu einem bestimmten Zweck kann zwar nicht sichergestellt werden. Wird jedoch der zum Abruf von Bewertungsdokumenten nötige Aufwand erhöht, so kann die Verwendung zum Zweck der Profilbildung zumindest erschwert werden. Der Schutz muss in diesem Fall durch die jeweiligen speichernden Knoten sichergestellt werden. Wie dies erreicht werden kann, wird im folgenden Abschnitt dargestellt.

5.5.2.3 Massenabruf von Dokumenten

Der massenhafte Abruf von Bewertungen kann – ebenso wie das massenhafte Einfügen – auf verschiedenen Wegen erschwert werden. Diesen liegt jedoch stets die Idee zugrunde, dem abrufenden Kosten zu verursachen, die bei einem einzelnen Abruf nicht ins Gewicht fallen, den systematischen Abruf zahlreicher Dokumente zur Profilbildung jedoch prohibitiv teuer werden lassen.[5]

Der einfache Ansatz, Antworten auf Dokumentanfragen zeitverzögert zu versenden, ist dabei aufgrund der möglichen Parallelisierung seitens des Angreifers nicht praktikabel. Stattdessen bietet sich die Verwendung sogenannter *crypto puzzles* an: Bevor eine Antwort versandt wird, muss der abfragende Knoten beweisen, dass er eine Rechenaufgabe gelöst hat. Die Lösung sollte dabei schwierig zu finden, aber leicht zu überprüfen sein. Ein entsprechender Ansatz wird bereits in [RiSW96] vorgestellt. Im Rahmen dieser Arbeit wird jedoch auf den in [RFC3972] und [CDGR+02] verfolgten Ansatz zurückgegriffen, eine kryptographische Hash-Funktion einzusetzen. Hierzu wird folgendes Verfahren verwendet:

Will ein Knoten die Bewertungsdokumente für ein Objekt abrufen, so kontaktiert er den für die Speicherung zuständigen Knoten. Die Anfrage enthält folgende Einträge:

- ID des verwendeten Objekts id.
- Zeitstempel t.
- Eine Zeichenkette S, für die gilt: Die letzten k bit von $hash(id|t|S)$ sind 0.

Der Parameter k dient dabei der Anpassung an im Laufe der Zeit steigende Rechenleistungen; $hash(x)$ ist eine kryptographische Hash-Funktion. Die Bedingung an die Zeichenkette S kann somit nur erfüllt werden, indem der anfragende Knoten so lange verschiedene Werte für S ausprobiert, bis die Bedingung erfüllt ist. Der antwortende Knoten prüft, ob er für das Objekt id zuständig ist, ob der Zeitstempel t aktuell – beispielsweise nicht älter als 5 Minuten – ist, und ob die Bedingung, dass die letzten k bit von $hash(id|t|S)$ den Wert 0 haben, erfüllt ist. Nur in diesem Fall wird die Anfrage beantwortet.

Ein Problem bei der Verwendung des Verfahrens liegt darin, dass eine Anpassung an steigende Rechenleistung über den Verlauf der Zeit hinweg möglich ist, nicht aber eine

[5] Einige Ansätze zur Abwehr des Sybil-Angriffs (vgl. Abschnitt 2.1.7) bauen auf einem ähnlichen Prinzip auf.

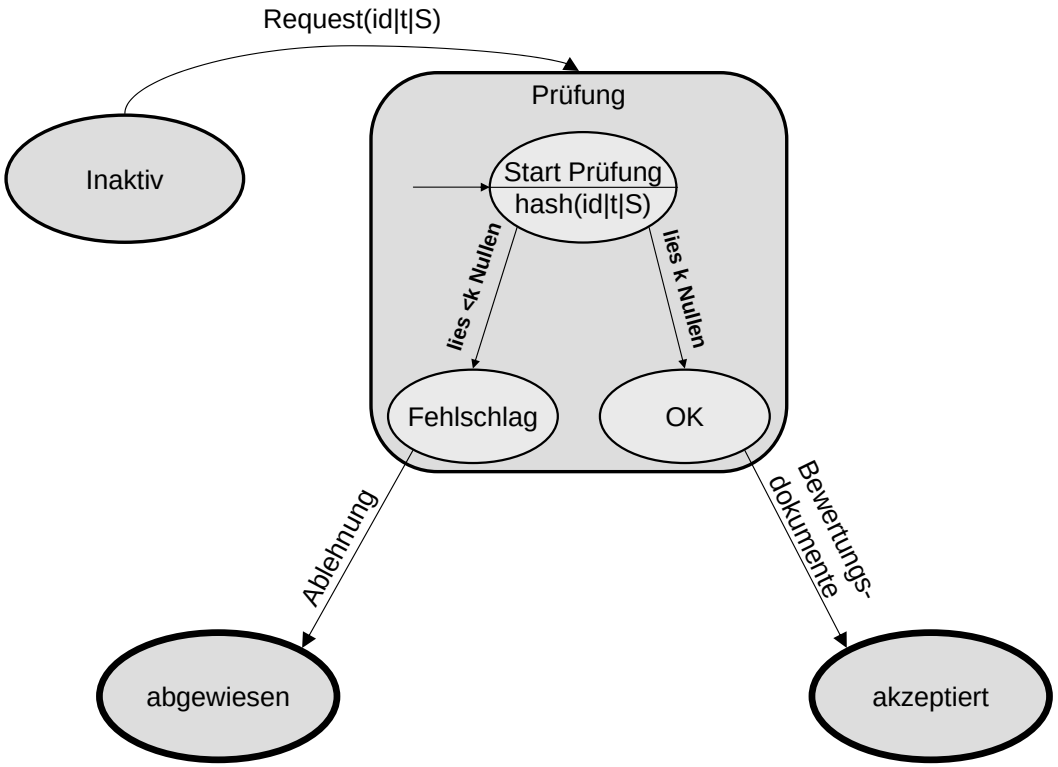

Abbildung 5.2: Abruf gespeicherter Bewertungen

Anpassung an die unterschiedlichen Rechenleistungen verwendeter Endgeräte. Es bietet sich deshalb an, die Konstante k an leistungsstarken Endgeräten auszurichten. Für Geräte mit geringerer Prozessorleistung – beispielsweise Mobiltelefone – soll eine kleinere Konstante k_m verwendet werden. In diesem Fall soll aber die Information über den Ersteller des Dokuments entfernt werden. Diese Lösung beeinträchtigt die Funktionalität des Systems, da der anfragende Knoten Bewertungen ihm bekannter Nutzer nicht mehr bevorzugen kann und zudem die Integrität und Authentizität übertragener Bewertungsdokumente sich nicht überprüfen lässt. Jedoch ist zumindest eine Teilfunktionalität auch für mobile Geräte noch vorhanden.

Der Abruf von Bewertungen ist in Abbildung 5.2 zusammengefasst. Der Knoten, der die Bewertung gespeichert hat – startet bei Erhalt einer Request-Nachricht (Anforderung einer Bewertung) den Prüfprozess zur Entscheidung, ob die Bewertung versendet werden kann. Dieser Prüfprozess ist im inneren Automaten dargestellt. Werden weniger als k Nullen gelesen, schlägt die Prüfung fehl; andernfalls ist sie erfolgreich, und der abrufende Knoten erhält die angeforderten Bewertungen. Für das Einfügen von Bewertungen findet das gleiche Verfahren Anwendung.

5.6 Nutzerbasiertes kollaboratives Filtern

Dieser Abschnitt beschreibt ein Verfahren zum nutzerbasierten Kollaborativen Filtern. Zunächst wird dabei ein grundlegendes nutzerbasiertes Verfahren entworfen, das im nächsten Schritt zu einem kombinierten Ansatz erweitert wird.

5.6.1 Basismodell

Bei der Verwendung nutzerbasierter Ansätze bestehen vier wesentliche Herausforderungen:

- Die Speicherung der Nutzerprofile.
- Das Ähnlichkeitsmaß, das durch diesen Algorithmus herangezogen wird.
- Der eigentliche Empfehlungsalgorithmus, der grundsätzlich beliebig gewählt werden kann.
- Der Algorithmus, der zum Finden ähnlicher Knoten (taste buddies) verwendet wird. Hierbei gilt es, einen Kompromiss aus Kommunikationsaufwand und der Auswahl möglichst passender Knoten zu finden.

5.6.2 Speicherung von Nutzerprofilen

Grundsätzlich kann bei einem nutzerbasierten Ansatz jeder Knoten sein eigenes Nutzerprofil speichern und auch selbst die Liste ähnlicher Knoten verwalten[6]. Um die Verfügbarkeit der Informationen auch dann zu gewährleisten, wenn ein Knoten nicht mit dem Peer-to-Peer-System verbunden ist, bietet sich auch hier die Verwendung der Datenspeicherungsschicht des Peer-to-Peer-Systems an. Als Schlüssel für die Speicherung wird die Identität des jeweiligen Nutzers verwendet.

- Hat ein Nutzer ein Objekt bewertet, so wird ein Bewertungsdokument erstellt und die Funktion *einfügen(Nutzer-ID, Bewertungsdokument)* des Datenspeicherungsdiensts aufgerufen.
- Soll eine Empfehlung für einen Nutzer erstellt werden, so wird für jeden taste buddy des suchenden Nutzers die Funktion `Dokument` *suchen(ID des taste buddy)* aufgerufen. Aus den so erhaltenen Dokumenten wird die Empfehlung berechnet.

Zusätzlich speichert jeder Knoten lokal das eigene Nutzerprofil sowie eine Liste der taste buddies mit deren Nutzer-IDs und Nutzerprofilen. Dazu wird eine Datenstruktur fester

[6] Bei einem objektbasierten Ansatz ist dies nicht möglich, da die Speicherung dort nicht nach Nutzern organisiert ist.

Größe t verwendet – könnte diese Datenstruktur beliebig wachsen, wäre die Skalierbarkeit des Verfahrens gefährdet. Sind alle Einträge dieser Datenstruktur belegt, werden weitere taste buddies nur dann hinzugefügt, wenn sie dem aktuellen Knoten ähnlicher sind als der bis dahin unähnlichste taste buddy; dieser wird in diesem Fall verdrängt. Hat ein Knoten seine taste-buddy-Liste aktualisiert, so speichert er sie unter Verwendung der Datenspeicherungsschicht mittels des Aufrufs *einfügen(Nutzer-ID, Liste der taste buddies)*.

5.6.3 Ähnlichkeitsmaß

Im Zusammenhang mit der vorliegenden Arbeit entstand eine erste Implementierung eines nutzerbasieren Ansatzes [Jäge06], die ähnlich [PSWS05] auf ein cosinus-basiertes Ähnlichkeitsmaß zurückgreift. Die Nutzerprofile werden dabei als Vektoren aufgefasst, die sich aus den Zeilen der in Abbildung 2.4 abgebildeten Matrix ergeben. Die Ähnlichkeit ist dann der Cosinus des Winkels zwischen diesen Vektoren:

$$Sim(\vec{A}, \vec{B}) = \cos(\vec{A}, \vec{B}) = \frac{\vec{A} \cdot \vec{B}}{|\vec{A}||\vec{B}|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{5.1}$$

Dieses Ähnlichkeitsmaß wird für die vorliegende Arbeit übernommen.

5.6.4 Empfehlungsalgorithmus

Der für die vorliegende Arbeit verwendete Empfehlungsalgorithmus gewichtet lediglich die Bewertungen der taste buddies mit deren Ähnlichkeitswert. Die vorhergesagte Bewertung, die einem Objekt O_A auf Knoten i zugewiesen wird, ergibt sich als

$$B_{O_A(i)} = \frac{\sum_{j \in tb(i), \exists B_{O_A(j)}} (Sim(i,j) \cdot B_{O_A(j)})}{\sum_{j \in tb(i), \exists B_{O_A(j)}} Sim(i,j)} \tag{5.2}$$

Es wird also für jeden Nutzer, der das jeweilige Objekt bewertet hat, das Produkt aus dessen Ähnlichkeit zum aktuell betrachteten Knoten mit der Bewertung des Objekts multipliziert. Die Ergebnisse werden aufsummiert und durch die Summe der Ähnlichkeiten der Nutzer, die das Objekt bewertet haben, geteilt. Da die vorhergesagte Bewertung einfach als gewichtetes Mittel gebildet wird, wird die Qualität der Vorhersage durch das als Gewicht verwendete Ähnlichkeitsmaß bestimmt. Die Datenstruktur, die die Bewertungen enthält, muss dabei den Zugriff auf Bewertungen anhand der Identität des Knotens, der sie abgegeben hat, ermöglichen. So kann auf vollständige Nutzerprofile zugegriffen werden. Konkret wird ein Vektor verwendet, der für jeden Knoten eine Hashtabelle enthält; als Schlüssel in diese Hashtabelle werden Objekt-IDs verwendet. Erforderlich ist auch die Möglichkeit, ohne Kenntnis der Schlüssel auf die gespeicherten Bewertungen zugreifen zu können, da dies für die effiziente Berechnung der Ähnlichkeit von Nutzerprofilen notwendig ist. Diese An-

forderung ist aber in der praktischen Implementierung einer Hashtabelle unproblematisch zu erfüllen.

Für die Aufgabe, eine Bewertung vorherzusagen, muss nun auf Knoten K (mit Nutzerprofil $\vec{K}$) lediglich der folgende Algorithmus durchgeführt werden, der als Eingabe die ID des Objekts O, dessen Bewertung vorherzusagen ist, sowie eine Menge von Bewertungen erhält.

1. Für jeden bewertenden Knoten X (mit Nutzerprofil $\vec{X}$), der O bewertet hat: Berechne $Sim(\vec{K}, \vec{X})$.

2. Berechne die Summe der in Schritt 1 berechneten Ähnlichkeiten.

3. Berechne $B_{O(K)}$ gemäß Gleichung 5.2.

Die Aufgabe „relevante Objekte finden“ wird wie folgt gelöst: Für jedes Objekt, das mindestens von einem taste buddy, aber nicht vom aktuellen Knoten bewertet wurde, wird die Operation „Bewertung vorhersagen“ ausgeführt. Die resultierenden Bewertungen werden von der besten Bewertung an absteigend sortiert und die ersten m Bewertungen dem Nutzer als relevante Objekte dargestellt.

5.6.5 Auffinden ähnlicher Knoten

Das Auffinden neuer taste buddies ist – aufbauend auf [Jäge06] – wie folgt umgesetzt:

- Zur Initialisierung kontaktiert jeder neue Netzteilnehmer eine gewisse Anzahl z' zufällig ausgewählter Knoten aus der eigenen Nachbarschaft. Als Nachbarschaft wird dabei die finger table des Chord-Verfahrens herangezogen, um sicherzustellen, dass nicht ausschließlich taste buddies aus der direkten Nachbarschaft gewählt werden. Die Anwendung greift dabei direkt auf die Overlay-Schicht zu, ohne die Datenspeicherungsschicht zu verwenden[7].

  Die Auswahl zufälliger Knoten durch einen Knoten x geschieht in folgender Weise:

  - Die Menge M der Nachbarn des Knotens ist gleich der Einträge in der finger table aus dem Chord-Protokoll. Sie hat m Elemente.
  - x berechnet $z = \lceil \frac{z'}{m} \rceil$
  - Ist $m \cdot z = z'$, so setzt x $z_i = z$ für $i \in 1, \ldots, m$. Ist $m \cdot z > z'$, so setzt x: $z_i = z - 1$ für $i \in 1, \ldots m \cdot z - z'$ und $z_i = z$, sonst.

[7] Zur Vereinfachung ist dieser – einzige – Fall der Umgehung der Datenspeicherungsschicht in Abbildung 5.1 nicht dargestellt.

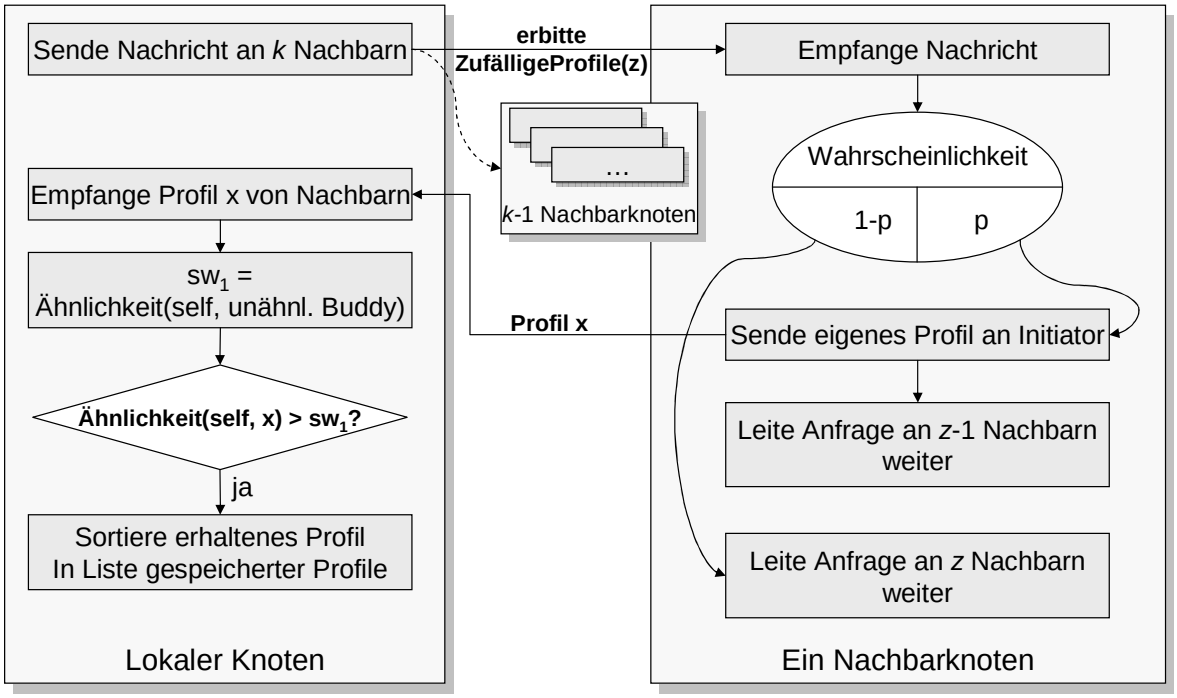

Abbildung 5.3: Finden ähnlicher Knoten

z_i muss also so gewählt werden, dass sich für alle kontaktierten Nachbarn i zusammen die Summe t (also die Gesamtzahl erwünschter taste buddies) ergibt. Nun sendet der Knoten eine Nachricht zum Aufbau einer Nachbarschaftsbeziehung an seine Nachbarn; in der Nachricht an Nachbar i ist dabei der Wert z_i enthalten. Mit Wahrscheinlichkeit p beantwortet der direkte Nachbar i die Nachricht selbst, sendet – falls ein solches vorhanden ist – sein gespeichertes Nutzerprofil an den anfragenden Knoten und leitet die Nachricht an $z - 1$ Knoten aus seiner eigenen Nachbarschaft weiter. Mit Wahrscheinlichkeit $1 - p$ leitet er die Nachricht an z seiner Nachbarn weiter und wird nicht zum taste buddy. Die Empfänger im nächsten Schritt verfahren nach derselben Methode. Das grundlegende Vorgehen bei diesem Protokoll wird in Abbildung 5.3 dargestellt.

- Jeder Knoten sendet periodisch Anfragen an die Knoten, die für die Speicherung der Nutzerprofile seiner taste buddies zuständig sind; dazu wird jeweils die Funktion *suchen(Nutzer-ID(taste buddy), Sucheinschränkung)* der Datenspeicherungsschicht genutzt. Die Sucheinschränkung besteht jeweils aus dem eigenen Nutzerprofil und dem Ähnlichkeitswert des unähnlichsten eigenen taste buddies (0, wenn die eigene taste-buddy-Liste nicht voll ist). Jeder Empfänger einer solchen Suchanfrage vergleicht dieses Nutzerprofil mit allen eigenen taste buddies. Findet er einen taste buddy, der dem

Nutzerprofil aus der empfangenen Nachricht ähnlicher ist als der in der Nachricht angegebene Ähnlichkeitswert, so antwortet er mit dessen Nutzerprofil.

5.6.6 Kombinierter Ansatz: Auffinden von Objekten

Eine große Schwäche des nutzerbasierten Verfahrens ist die eingeschränkte Möglichkeit, Objekte aufzufinden. Möchte ein Nutzer eine personalisierte Bewertung eines bestimmten Objekts, so kann er diese nur erhalten, wenn mindestens einer seiner taste buddies dieses Objekt bewertet hat. Die Bewertungen anderer Knoten im Netz für dieses Objekt kennt der Knoten jedoch nicht. Um diesen Mangel zu beheben, ohne jedoch die Vorteile des nutzerbasierten Verfahrens aufzugeben, wird folgendes Verfahren vorgeschlagen:

- Hat ein Knoten seine taste buddies gefunden, so prüft er für jedes Objekt aus seinem Nutzerprofil, ob dieses mindestens von γ seiner taste buddies bewertet wurde. Die selbe Überprüfung wird auch durchgeführt, wenn der Knoten ein neues Objekt bewertet.
- Ist dies nicht der Fall, so speichert er einen Verweis auf seine eigene Identität unter dem Schlüssel des entsprechenden Objekts (Aufruf von *einfügen(Schlüssel(Objekt), eigene Identität)*).
- Der oder die für diesen Schlüssel zuständigen Knoten speichern bis zu η Identitäten von Nutzern, die das Objekt bewertet haben; die ältesten Bewertungen werden verdrängt.
- Sucht nun ein Knoten eine Bewertung eines Objekts, das seine taste buddies nicht bewertet haben, so sendet er eine Anfrage an den Knoten, der für das entsprechende Objekt zuständig ist, und fügt die dort gespeicherten Nutzerprofile temporär seinen taste buddies hinzu. Hierbei werden keine bestehenden taste buddies verdrängt; die Speicherung erfolgt zusätzlich zu den vorhandenen Nutzerprofilen.

Auf diesem Weg kann die Anzahl an Objekten, für die eine Bewertung vorhergesagt werden kann, deutlich erhöht werden. Dazu werden Elemente eines objektbasierten Ansatzes eingeführt, jedoch nur für diejenigen Objekte, die mit Hilfe des nutzerbasierten Ansatzes wahrscheinlich nicht gefunden werden könnten. Es entsteht hier also eine Kombination aus nutzer- und objektbasiertem Ansatz.

5.7 Verteilte Speicherung von Bewertungsdokumenten

In diesem Abschnitt wird die verteilte Speicherung von Bewertungsdokumenten dargestellt. Im Grundsatz wird ein Bewertungsdokument dabei lediglich über die Funktion *einfü-*

gen(Schlüssel, Dokument) der Datenspeicherungsschicht eingefügt. Diese Funktionalität kann jedoch noch ergänzt werden.

5.7.1 Reputation und Schutz von Identitäten

Zwischen der Funktionalität eines Reputationssystems und dem Schutz der Identität von Systemteilnehmern besteht ein Spannungsverhältnis: Reputation ist stets mit einem Träger – üblicherweise einer natürlichen oder juristischen Person – verknüpft, was jedoch zum Vorliegen personenbezogener Daten führen kann.

Als Lösungsmöglichkeit bietet sich zum einen an, statt der Nutzer nur einzelne Bewertungen zu bewerten. Diese Möglichkeit nimmt den Teilnehmern den Wohlverhaltensanreiz, wie er durch Bewertungen der Nutzer selbst erreicht werden kann; die Möglichkeit, Bewertungen zu bewerten, wird dennoch vorgesehen.

Zum anderen kann jedoch auch die Verwendung von Pseudonymen Abhilfe bringen. Dabei ist zu beachten, ein Nutzer, der einen schlechten Ruf erlangt hat, diesen nicht einfach durch Wechsel des Pseudonyms verbessern können soll. Zunächst werden daher auf Anwendungsebene die Identitäten der Knoten aus der Overlay-Ebene übernommen, so dass die Maßnahmen zur Abwehr des Sybil-Angriffs (vgl. Seite 18) auch das Reputationssystem schützen.

Zusätzlich könnte der Pseudonymwechsel auch auf Anwendungsebene mit Kosten verbunden werden. Diese müssen nicht monetärer Natur sein; denkbar ist auch

- von einem Pseudonymträger zu verlangen oder mit erhöhter Reputation zu belohnen, dass dieser einen Beweis für investierte Ressourcen (wie Rechenaufwand) erbringt, falls dies nicht bereits auf Overlay-Ebene geschieht – ein Ansatz, der beispielsweise in [RFC3972] verfolgt wird –,
- den Anreiz zum Pseudonymwechsel zu nehmen, indem neu erstellte Pseudonyme mit der schlechtestmöglichen Reputation ausgestattet werden – die Kosten des Pseudonymwechsels liegen in diesem Fall in der verlorenen Reputation.

Im Rahmen des entwickelten Systems wird darauf jedoch verzichtet und die Verwaltung von Identitäten allein auf der Overlay-Ebene vorgenommen. Die zitierten Verfahren sind in der Literatur bereits hinreichend diskutiert und können dem System bei Bedarf hinzugefügt werden.

Wie kann der Einsatz eines Reputationssystems aber Fehlverhalten von Teilnehmern reduzieren? Dies setzt zum einen voraus, dass Fehlverhalten eines Teilnehmers entdeckt und mit einer Verschlechterung seiner Reputation geahndet werden kann; zum anderen muss ein schlechter Ruf auch zu negativen Konsequenzen für den jeweiligen Teilnehmer führen.

Die erste Voraussetzung ist ohne Weiteres zu erfüllen. Aufgrund der eingefügten Bewertungsdokumente können andere Peers Nutzer bewerten. Zu bedenken gilt jedoch, dass eine

schlechte Reputation in einem Empfehlungssystem nicht notwendigerweise zu Nachteilen führt. Jedoch könnte eine solche zum Anlass genommen werden,

- Bewertungsdokumente des jeweiligen Nutzers nur mit geringem Gewicht zu berücksichtigen. Dies nimmt den Anreiz, zur Förderung bestimmter Objekte (beispielsweise eigener Produkte) in großer Zahl Bewertungsdokumente zur Verfügung zu stellen. Wie dies konkret umgesetzt wird, wird in Abschnitt 5.7.2 gezeigt.
- Anfragen des entsprechenden Knotens niedriger zu priorisieren bzw. Antworten zu verzögern, wenn dieser selbst Bewertungsdokumente abrufen will. Dies betrifft normale Nutzer des Systems, die Bewertungen schlechter Qualität abgeben.

Die Verwendung solch dauerhafter Pseudonyme birgt jedoch auch Risiken. Gelingt es einem Angreifer, Bewertungsdokumente zu sammeln, die durch einen einzelnen Knoten erstellt wurden, so besteht die Gefahr, dass ein genaues Profil über den Pseudonymträger erstellt werden kann – je nach Art der bewerteten Objekte kann so ein Pseudonym auch aufgedeckt werden (vgl. zu den möglichen Angriffen auch [RKMG+01]).

Die Erstellung eines solchen Profils wird daher durch den in Abschnitt 5.5.2.3 beschriebenen Mechanismus schon in der Datenspeicherungsschicht erschwert. Durch das Verfahren wird sichergestellt, dass beim Abruf eines Bewertungsdokuments Kosten (in Form von Rechenaufwand) entstehen; das Erstellen von Nutzerprofilen wird somit nicht unmöglich, aber wenig attraktiv.

5.7.2 Bewertungen von Bewertungen

Zusätzlich zur Bewertung von Objekten kann Information zur Verfügung gestellt werden, die die Beurteilung der Bewertungen erlaubt. Hierzu werden Bewertungen in der gleichen Weise bewertet wie Objekte selbst. Weitere Bewertungsebenen sind nicht vorgesehen: Die Bewertung einer Bewertung kann nicht bewertet werden – dem dadurch entstehenden geringen Nutzen stünde das Risiko eines DoS-Angriffs durch Einfügen beliebig vieler Bewertungen immer tieferer Ebenen gegenüber.

Für die Bewertungen von Bewertungen wird kein neuer Schlüssel definiert, der zur Zuständigkeit eines anderen Knotens führen könnte. Stattdessen wird der Schlüssel des zugehörigen Objekts verwendet, sodass die gleichen Knoten, die für Bewertungen des Objekts zuständig sind, auch die Bewertungen der Bewertungen speichern.

Die Bewertungen von Bewertungen werden im Nutzer-Interface herangezogen, um die Reihenfolge zu steuern, in der Objektbewertungen angezeigt werden, und die Aggregation von Bewertungszahlen zu einer Gesamtbewertung zu ermöglichen. Sind p der Identifikator eines Objekts, b_p^i die Bewertungsdokumente zu diesem Objekt, $a(b_p^i)$ (mit $i = 1, \ldots,$ Anzahl der Bewertungsdokumente zu Objekt p) die zugehörigen Bewertungszahlen, $id(b_p^i)$

die Identifikatoren der jeweiligen Bewertungsdokumente, k_i die Anzahl von Bewertungen der Objektbewertung i sowie $t(author(b_p^i))$ das Vertrauen in deren Autor, so ergibt sich das Gewicht dieser Objektbewertung $w(b_p^i)$ als

$$w(b_p^i) = \frac{1}{2} \cdot t(author(b_p^i)) + \frac{1}{2k_i} \sum_{j=1}^{k_i} a(b^j_{id(b_p^i)}) \cdot t(author(b^j_{id(b_p^i)})) \tag{5.3}$$

Somit wird sowohl die Sicht des jeweiligen Nutzers (durch die Einbeziehung seines Vertrauens in den Autor) als auch die anderer Nutzer im System (aufgrund ihrer Bewertungen) berücksichtigt. Der Vertrauenswert ergibt sich durch die Verwendung eines Reputationssystems oder aufgrund manueller Auswahl durch den Nutzer. Werden keine Vertrauenswerte vergeben, wird der entsprechende Parameter auf den Wert 1 gesetzt; sie könnte jedoch auch als Funktion der Anzahl durch den jeweiligen Autor abgegebener Bewertungen bestimmt werden.

Zur Aggregation der Objektbewertungen über ihren Zahlenwert wird die Gesamtbewertung eines Objekts gebildet als

$$a_p^{agg} = \sum_{i=1}^{l} w(b_p^i) \cdot a(b_p^i), \tag{5.4}$$

wobei l die Gesamtzahl der Bewertungen des Objekts ist. Dies bedeutet eine einfache Gewichtung der eingehenden Einzelbewertungen.

5.8 Objektbasiertes Kollaboratives Filtern

Die verteilte Speicherung von Bewertungsdokumenten ermöglicht zwar das Auffinden von Informationen über Objekte; sie erlaubt es aber noch nicht, Empfehlungen im Sinne einer individuellen Vorhersage der Bewertung von Objekten oder des Auffindens relevanter Objekte zu erstellen. Im Folgenden soll nun ein objektbasierter Ansatz des Kollaborativen Filterns dargestellt werden, der eine Alternative zum bereits vorgestellten nutzerbasierten Ansatz darstellt.

5.8.1 Basismodell

Die grundlegende Idee des Verfahrens besteht darin, in einer DHT jeweils unter dem Schlüssel eines Objekts die Information abzulegen, welche anderen Objekte mit diesem (wie oft) gemeinsam benutzt wurden (vgl. Abb. 5.4). Eine gemeinsame Benutzung ist dabei als das Vorhandensein der Objekte in einem Nutzerprofil definiert. Die verwendete Datenstruktur heißt auch *Objekttabelle*. Im Gegensatz zum nutzerbasierten Ansatz werden also nicht die Zeilen, sondern die Spalten der Matrix aus Abbildung 2.4 jeweils unter einem gemeinsamen Schlüssel gespeichert. Eine Objekttabelle enthält konkret die folgenden Einträge:

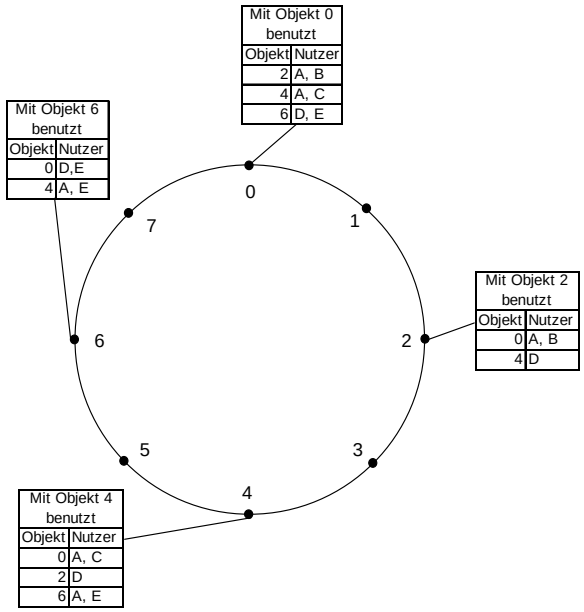

Abbildung 5.4: Speicherung von Objekttabellen in Chord

- Gesamtzahl der Benutzungen des Objekts
- Für jede gemeinsame Nutzung mit einem anderen Objekt (das heißt, für jedes Vorhandensein beider Objekte im Nutzerprofil eines Nutzers):
  - Identifikator des anderen Objekts
  - Zeitpunkte der Benutzung beider Objekte
  - Bewertungszahlen beider Objekte (wie durch den Nutzer vergeben)
  - Identität (Pseudonym) der benutzenden Person
  - Signatur der benutzenden Person

Die tatsächliche Einbeziehung der Bewertungszahlen durch den verwendeten Empfehlungsalgorithmus ist dabei nicht nötig; dies hängt vom jeweiligen Anwendungsszenario ab. Der Empfehlungsalgorithmus ist dabei leicht austauschbar.

Jeder Knoten hat zudem eine Liste aller Objekte, die der jeweilige Nutzer bereits in Benutzung hatte, inklusive der zugehörigen Bewertungsdokumente (Nutzerprofil). Wünscht der Nutzer nun eine Empfehlung, so wird bei den für jedes lokal bewertete Objekt jeweils zuständigen Knoten angefragt, welche anderen Obekte mit diesem gemeinsam benutzt wurden. Aus dieser Information kann dann lokal eine Empfehlung erzeugt werden. Benutzt ein Teilnehmer ein weiteres Objekt, so muss er alle Knoten benachrichtigen, die für ein von ihm genutztes Objekt zuständig sind, d.h. er muss die Information über die gemeinsame Benutzung unter den Schlüsseln dieser Objekte speichern. Dies führt auch schon zum bereits angesprochenen Skalierbarkeitsproblem: Der benötigte Kommunikationsaufwand für

die Aktualisierung (bei Erwerb eines neuen Objekts) liegt bei einer Nutzerprofil-Größe von u in $O(u)$, betrachtet man die Anzahl der Nachrichten auf der Overlay-Ebene. In der Transportschicht ergeben sich daraus $O(u \cdot \log n)$ Nachrichten (n Anzahl der Knoten im Overlay) (vgl. [SMKK+01]). Für den Aufbau aller Objekttabellen – unterstellend, dass immer ein Objekt auf einmal den Tabellen hinzugefügt wird – ergibt sich gar ein quadratischer Aufwand. Zusammengefasst für alle m Nutzer des Systems sind auf Overlay-Schicht $O(m \cdot u^2)$ Nachrichten und auf Transportschicht $O(m \cdot u^2 \cdot \log(n))$ Nachrichten zu versenden. Wird vereinfachend $m = n$ gesetzt (jedem Nutzer also genau ein Knoten zugeordnet), sind es $O(n \cdot u^2)$ Nachrichten auf Overlay- und $O(n \cdot u^2 \cdot \log(n))$ auf Transportschicht. Will ein Knoten aktualisierte Objekttabellen abrufen, so sind dazu ebenfalls $O(u)$ Nachrichten auf Anwendungs- bzw. $O(u \cdot \log n)$ Nachrichten auf Transportschicht erforderlich. Unterstellt man, zu jedem erworbenen Objekt sei vorher eine solche Aktualisierung vorgenommen worden, ergibt sich hierfür der gleiche Gesamtaufwand wie bereits für das Versenden der eigenen Aktualisierungsnachrichten.

Im Folgenden werden zunächst funktionale Verbesserungen betrachtet, die die Skalierbarkeit erheblich verbessern und eine hohe Empfehlungsqualität ermöglichen. Anschließend werden Sicherheitsmechanismen eingeführt, die auch die Anbindung eines Reputationssystems umfassen und den Schutz personenbezogener Daten im System ermöglichen.

5.8.2 Empfehlungsberechnung

Der verwendete Empfehlungsalgorithmus basiert auf der Häufigkeit gemeinsamer Benutzungen zweier Objekte. Er lehnt sich an einen in [DeKa04] vorgeschlagenen Algorithmus an. Für alle Objekte, die für den Nutzer interessant sein könnten – also alle, die mindestens einmal gemeinsam mit einem Objekt aus seinem Nutzerprofil verwendet wurden –, wird die *Relevanz* berechnet. Dazu wird folgende Berechnung zugrunde gelegt:

$$R(O_B, O_A) = \frac{F_\cap(O_A, O_B)}{F(O_A) \cdot F(O_B)^\alpha}, 0 \leq \alpha \leq 1 \tag{5.5}$$

$$NR_i(O_B, O_A) = \frac{R(O_B, O_A)}{\sum_{O_K \in up_i} R(O_K, O_A)} \tag{5.6}$$

mit $\alpha \in [0, 1]$ Konstante. Die Relevanz eines Objekts B für ein Objekt A ergibt sich also durch Division der Anzahl gemeinsamer Benutzungen $F_\cap$, geteilt durch das Produkt aus der Anzahl an Benutzungen von Objekt A und der mit α potenzierten Benutzungshäufigkeit von Objekt B. Die normierte Objekt-zu-Objekt-Relevanz wird aus Sicht eines Nutzers i über alle in seinem Nutzerprofil up_i enthaltenen Objekte gebildet.

Sei nun $E_z(X)$ die Bewertung, die Nutzer z einem Objekt X gegeben hat und

$$\overline{E}_{X,Y}(X) = \frac{\sum_{z|X \in up_z, Y \in up_z} E_z(X)}{\sum_{z|X \in up_z, Y \in up_z} 1} \tag{5.7}$$

die durchschnittliche Bewertung für das zu bewertende Objekt durch Nutzer, die auch das jeweilige andere Objekt bewertet haben.

Die vorhergesagte Bewertung E_i^* ergibt sich dann durch eine einfache Gewichtung

$$E_i^*(O_A) = \sum_{O_B \in up_i} NR_i(O_B, O_A) \cdot \overline{E}_{O_A,O_B}(O_A) \tag{5.8}$$

Ausgehend von diesem Berechnungsverfahren lässt sich eine Rangordnung der für einen Nutzer relevantesten Objekte, die ihm vorher nicht bekannt waren, erstellen, oder auch eine Bewertung eines einzelnen Objekts vorhersagen. Es ist denkbar, dass das Berechnungsverfahren keine Bewertungsvorhersage erstellen lässt – beispielsweise, wenn das Nutzerprofil des berechnenden Knoten i leer ist oder keines der darin enthaltenen Objekte gemeinsam mit dem Zielobjekt O_A benutzt wurde. In diesem Fall wird die durchschnittliche Bewertung von O_A als Ergebnis zurückgeliefert.

Die nächsten Abschnitte sollen zeigen, wie basierend auf dem dargestellten Berechnungsverfahren Empfehlungen in einem Peer-to-Peer-System erzeugt werden können.

5.8.2.1 Ablauf der Empfehlungserzeugung

Die Empfehlungserzeugung kann grundsätzlich auf zwei verschiedenen Wegen stattfinden

- Im ersten Fall sollen dem Nutzer aufgrund seines Profils für ihn relevante weitere Objekte empfohlen werden („Finde relevante Objekte“). Die Empfehlungserzeugung läuft dabei wie folgt ab:
  - Der Knoten ruft die Objekttabellen zu den in seinem Nutzerprofil vorhandenen Objekten ab bzw. fordert eine Aktualisierung an. Er führt also die Operation *suchen(Schlüssel)* für jedes dieser Objekte aus.
  - Zu den damit gemeinsam genutzten Objekten schätzt er aufgrund der nun vorhandenen Informationen deren Relevanz ab und fordert von den k relevantesten wiederum die Objekttabellen an. Zur Abschätzung, welches die relevantesten Objekte sind, wird in Gleichung 5.5 die Nutzungshäufigkeit F des nicht im lokal vorhandenen Nutzerprofil enthaltenen Objekts auf 1 gesetzt, da die Information über die tatsächliche Nutzungshäufigkeit noch nicht vorhanden ist. Die geschätzte Relevanz ist dann

$$R^*(O_B, O_A) = \frac{F_\cap(O_A, O_B)}{F(O_B)^\alpha} \tag{5.9}$$

  - Aus den nun vorhandenen, normierten Informationen wird aufgrund der aus Gleichung 5.6 errechneten Relevanzen eine Rangliste der relevantesten, im lokalen Nutzerprofil noch nicht vorhandenen Objekte erstellt. Die obersten Einträge werden dem Nutzer empfohlen (Anzahl je nach Auswahl des Nutzers).

- Im zweiten Fall soll eine nutzerspezifische Bewertung – also eine Prognose über die Einschätzung eines Objekts durch einen Nutzer – erzeugt werden. Dazu wird die Objekttabelle des jeweiligen Objekts angefordert und dessen Bewertung gemäß Gleichung 5.8 berechnet.

Im Vergleich zum naiven Verfahren, bei dem für jeden Schritt eine Nachricht an alle Knoten gesendet wird, die die Objekttabelle des jeweils gewünschten Objekts verwalten, sind diverse Verbesserungen möglich.

5.8.3 Funktionale Optimierungen

Den ersten Schritt bilden funktionale Optimierungen, die einen Teil der benötigten Nachrichten einsparen können. Hierzu werden verschiedene Ansätze verfolgt:

- Auf der Ebene der Datenhaltung:
  - Die Replikation von Objekttabellen in verschiedenen Bereichen des Chord-Rings.
  - Die Verwendung eines einfachen Anwendungs-Multicast-Protokolls zum Versenden von Aktualisierungsnachrichten.

- Auf der Ebene der Empfehlungsanwendung:
  - Die Reduktion der Häufigkeit von Aktualisierungen von Objekttabellen.
  - Die Speicherung der Objekttabellen verschiedener Objekte an einem gemeinsamen Speicherort (Schlüssel).
  - Die Zusammenfassung mehrerer Aktualisierungsnachrichten verschiedener Objekte.

Diese Verbesserungen werden im Folgenden beschrieben.

5.8.3.1 Replikation von Objekttabellen

Sollte eine verbesserte Lastverteilung nötig werden so bietet es sich an, gespeicherte Inhalte auch auf andere Knoten zu replizieren. Durch die Besonderheiten des Chord-Routings kann dabei gleichzeitig der Kommunikationsaufwand auf Kosten des Speicheraufwands verringert werden. Hierzu sind verschiedene Verfahren denkbar:

- Der in [DBKK+01] vorgeschlagene Caching-Mechanismus (vgl. Abschnitt 5.5) könnte eingesetzt werden.

- Alternativ soll an dieser Stelle ein durch den speichernden Knoten initiierter Replikationsmechanismus vorgestellt werden. Ziel ist dabei, diesem Knoten die Kontrolle über die Replikation zu geben, um im Bedarfsfall eine schnellere Aktualisierung der repliziert gespeicherten Daten herbeiführen zu können.

  Der Chord-Ring wird in m Teile unterteilt (mit $m \in \{1, 2, 4, \ldots\}$). Jeder Knoten bestimmt m dabei in Abhängigkeit der Netzlast im Overlay. Grundsätzlich sind beliebige Funktionen zur Ermittlung der Netzlast anwendbar. Notwendig ist lediglich ihre Nachvollziehbarkeit. Für die Evaluierung des Verfahrens wird die lokale Knotendichte ermittelt und als Stellvertreter für die Netzlast herangezogen. Somit ergibt sich für den Parameter m

$$m = 2^{\lceil l \cdot \frac{k}{d(cid_o, cid_s)} \rceil} \tag{5.10}$$

  mit l Konstante, cid_o eigene Chord-ID, cid_s Chord-Id des Nachfolgeknotens, $d(x, y)$ Distanz der Knoten x und y im Chord-Ring (als Abstand der Chord-IDs ermittelt), k Größe des Wertebereichs der für Chord verwendeten Hashfunktion.

  Im Normalfall soll – abgesehen von der auf der Datenspeicherungsschicht nötigen Redundanz – nach wie vor nur ein einzelner Knoten für einen Schlüssel verantwortlich sein. Steigt dessen Belastung, werden jedoch zusätzliche zuständige Knoten bestimmt. Dies geschieht dadurch, dass einem Objekt zusätzliche Speicherorte zugewiesen werden. Entscheidet ein Knoten sich, die bei ihm vorhandenen Daten zu replizieren, so wird der für m ermittelte Wert aber lediglich als obere Grenze verwendet; der Knoten darf die Replikation bei dieser Anzahl anderer Knoten vornehmen, muss aber nicht. Diese Entscheidung wird in Abhängigkeit der Auslastung des Knotens auf Anwendungsebene gefällt.

$$ort_j = ort_1 + \left\lceil \frac{(j-1) \cdot k}{m} \right\rceil \bmod k, \tag{5.11}$$

wobei $j = 1, \ldots, m$, k die Größe des Wertebereichs der verwendeten Hashfunktion und ort_1 der im ursprünglichen Chord-Protokoll zugewiesene Speicherort (also hash(Schlüssel)) ist. Der für diesen Schlüssel zuständige Knoten kann entscheiden, seine Objekttabelle an lokal – für einen bestimmten Bereich des Chord-Rings – zuständige Knoten zu senden, sodass sie insgesamt bis zu m mal gespeichert sind. Durch Beobachtung der jeweils lokal vorliegenden Knotendichte kommen andere Knoten zu einem ähnlichen m-Wert und können den jeweils nächstgelegenen Knoten kontaktieren, der die gewünschte Objekttabelle repliziert. Liegt die Objekttabelle dort nicht vor, so leitet dieser Knoten sie an den nächsten zuständigen Knoten

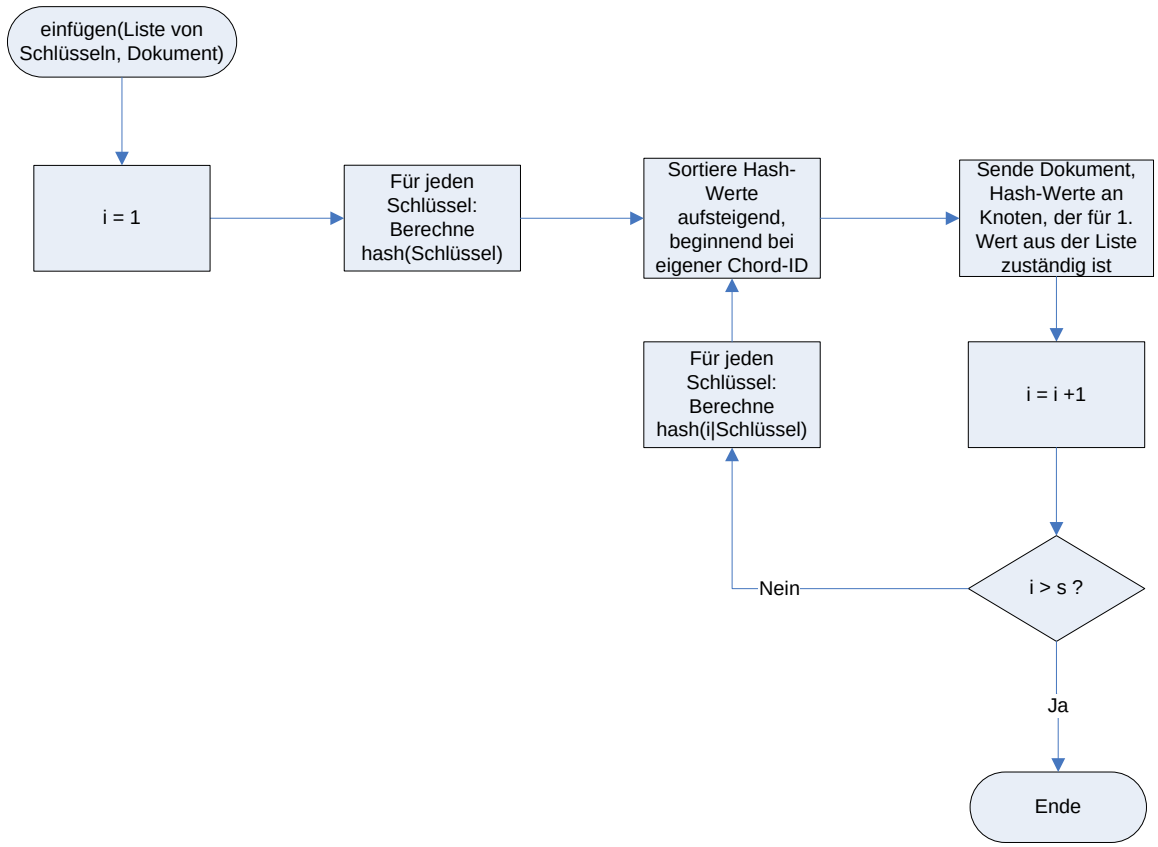

Abbildung 5.5: Multicast in der Datenspeicherungsschicht

weiter, der für den sich aus Gleichung 5.11 ergebenden Speicherort unter Verwendung des halbierten Werts für m zuständig ist.

Vorteil dieses Verfahrens ist, dass der zuständige Knoten die Replikation selbst kontrollieren und dementsprechend auch bei Aktualisierungen schnell reagieren kann. So kann die Konsistenz des Datenbestands weitgehend gewährleistet werden.

5.8.3.2 Multicast auf Anwendungs-Ebene

Betrachtet man die von einem Knoten sowohl bei der Aktualisierung der lokal gespeicherten Empfehlungsdaten als auch nach der Benutzung eines Objekts versandten Nachrichten, so fällt auf, dass fast identischer Nachrichteninhalt an eine Vielzahl anderer Knoten im Chord-Ring gesendet wird: Im einen Fall besteht die Nachricht lediglich aus der Anfrage nach der Zusendung einer aktualisierten Objekttabelle; im anderen Fall wird das Benutzungsprofil des jeweiligen Knotens versandt. Es bietet sich an, die erforderliche Nachrichtenzahl durch Verwendung eines Multicast-Protokolls auf der Overlay-Ebene – wie sie in allgemeiner Form in [Hueb03] beschrieben sind – zu reduzieren. Die Datenspeicherungsschicht wird dazu um eine Methode *einfügen(Liste von Schlüsseln, Dokument)* ergänzt. Hier-

zu wird eine einfache Variante gewählt: Der Absender bestimmt die Zielschlüssel, sortiert die zugehörigen Speicherorte in aufsteigender Reihenfolge – beginnend mit seiner eigenen Knoten-ID – und sendet sie dann an den ersten Knoten in der Liste. Jeder Empfänger entfernt sich von der Empfängerliste und leitet die Nachricht an den ersten folgenden Speicherort auf der Liste weiter. Da die durchschnittliche Entfernung zum Folgeknoten geringer ist, als wenn der Ausgangsknoten seine Nachricht an jeden einzelnen Empfänger senden müsste, wird die Anzahl erforderlicher Hops auf diesem Wege reduziert. Um weiterhin die Redundanz auf der Datenspeicherungsebene zu ermöglichen, wird das Protokoll einzeln für jeweils ein Replikat ausgeführt. Abbildung 5.5 verdeutlicht dieses Verfahren aus Sicht des ursprünglich einfügenden Knotens: Ist der Parameter s aus Abschnitt 5.5 gleich 1, so ist das Verfahren nach Sortieren der Speicherorte und Versenden an den ersten Speicherort der Liste abgeschlossen; andernfalls wird es für jede Redundanzebene mit den dann jeweils gültigen Speicherorten wiederholt.

5.8.3.3 Aggregierte Aktualisierung

Ein erster Schritt zur Verringerung des Kommunikationsaufwands besteht darin, Aktualisierungen von Objekttabellen nicht bei jeder Änderung vorzunehmen, sondern stets mehrere Aktualisierungen zusammenzufassen.

Für den *Abruf* von aktualisierten Objekttabellen können folgende Grundsätze herangezogen werden:

- Solange wenige Einträge vorhanden sind, sollten Aktualisierungen häufiger vorgenommen werden; spätere Aktualisierungen können mit größerer Verzögerung vorgenommen werden . In [Jäge06] wird dazu vorgeschlagen, jeweils erst nach $2^i (i \in \{1, \ldots\})$ lokalen Aktualisierungen eine Änderung zu propagieren. Dieser Vorschlag bietet einen Ansatzpunkt, doch führt seine Implementierung bei großen Nutzerprofilen dazu, dass praktisch gar keine Aktualisierungen mehr vorgenommen werden. Im Rahmen dieser Arbeit wird die Aktualisierungsfrequenz daher nicht exponentiell reduziert, sondern lediglich linear.

- Einträge, die nur mit geringem Gewicht in die Empfehlungsberechnung einfließen, erfordern seltener eine Aktualisierung. Das setzt jedoch voraus, dass dem Knoten, der die Aktualisierung anstößt, diese Einträge bekannt sind – funktioniert also nicht während der Initialisierung eines Knotens. Zur Unterstützung dieser Funktion lässt sich die Relevanz abschätzen, mit der das entsprechende Objekt O_A in die Empfehlung für einen Nutzer eingeht. Als Schätzfunktion wird verwendet:

$$ER(O_A) = F(O_A) \tag{5.12}$$

Die geschätzte Relevanz ER des Objekts entspricht also seiner Häufigkeit. Dementsprechend kann auch der Abruf der Objekttabelle selten benutzter Objekte mit geringer Frequenz geschehen. Dies gilt zumindest für das Empfehlen dem Nutzer noch nicht bekannter Objekte; die Bewertung eines Objekts vorherzusagen, erfordert ohnehin einen deutlich geringeren Aufwand.

Zusammenfassend lässt sich die Dauer zwischen zwei aufeinanderfolgenden Aktualisierungen der lokalen Kopie der Objekttabelle für Objekt O_A wie folgt beschreiben: Die Aktualisierung findet statt, wenn eine der beiden folgenden Bedingungen erfüllt ist.

$$\Delta t = c \cdot \frac{|E|}{F(O_A)} \tag{5.13}$$

$$\Delta E = d \cdot \frac{|E|}{F(O_A)} \tag{5.14}$$

wobei $|E|$ die Gesamtzahl durch den Nutzer angeforderter Empfehlungen ist; c und d sind Konstanten. Die konkrete Umsetzung ist in Abbildung 5.6 beschrieben. Jeder Knoten speichert für alle Objekte, deren Objekttabellen vorhanden sind, jeweils die Anzahl an Schritten bis zur nächsten Aktualisierung (in Abbildung 5.6: „Schritte bis Aktualisierung“) sowie den spätesten Zeitpunkt (in Abbildung 5.6: „Zeit für Aktualisierung“) der nächsten Aktualisierung. Als „Schritt“ wird dabei jede Verwendung der Objekttabelle durch den Empfehlungsalgorithmus gezählt.

Für jedes Objekt, dessen Objekttabelle durch den Empfehlungsalgorithmus benötigt wird, prüft das Verfahren, ob diese Tabelle bereits lokal vorhanden ist. Ist dies nicht der Fall, wird die Tabelle angefordert. Ansonsten wird der Schrittzähler für das Objekt geprüft; ist er bei 0 angelangt, wird die Objekttabelle angefordert, sonst der Zähler dekrementiert. Die Objekttabelle wird auch angefordert, wenn der vorgegebene Zeitpunkt für eine Aktualisierung verstrichen ist.

Das Anfordern einer Objekttabelle wird durch die Operation suchen(Schlüssel(Objekt)) der Datenspeicherungsschicht realisiert. Nach dieser Anforderung werden Schrittzähler und spätester Zeitpunkt der nächsten Aktualisierung für das jeweilige Objekt gemäß Gleichung 5.14 neu gesetzt.

Für das *Einfügen* neuer Objektbewertungen kann differenziert werden:

- Den Objekttabellen der neu hinzugekommenen Objekte müssen die bereits vor der neu hinzugefügten Bewertung im lokalen Nutzerprofil enthaltenen Objekte hinzugefügt werden. Diese Vorgang ist als solcher für jedes neue Objekt einmalig; hier können zwar mehrere neue Objekte zusammengefasst werden, doch eine Priorisierung nach Objekten ist nicht nötig.

- Den Objekttabellen der bereits vor der neu hinzugefügten Bewertung im lokalen Nut-

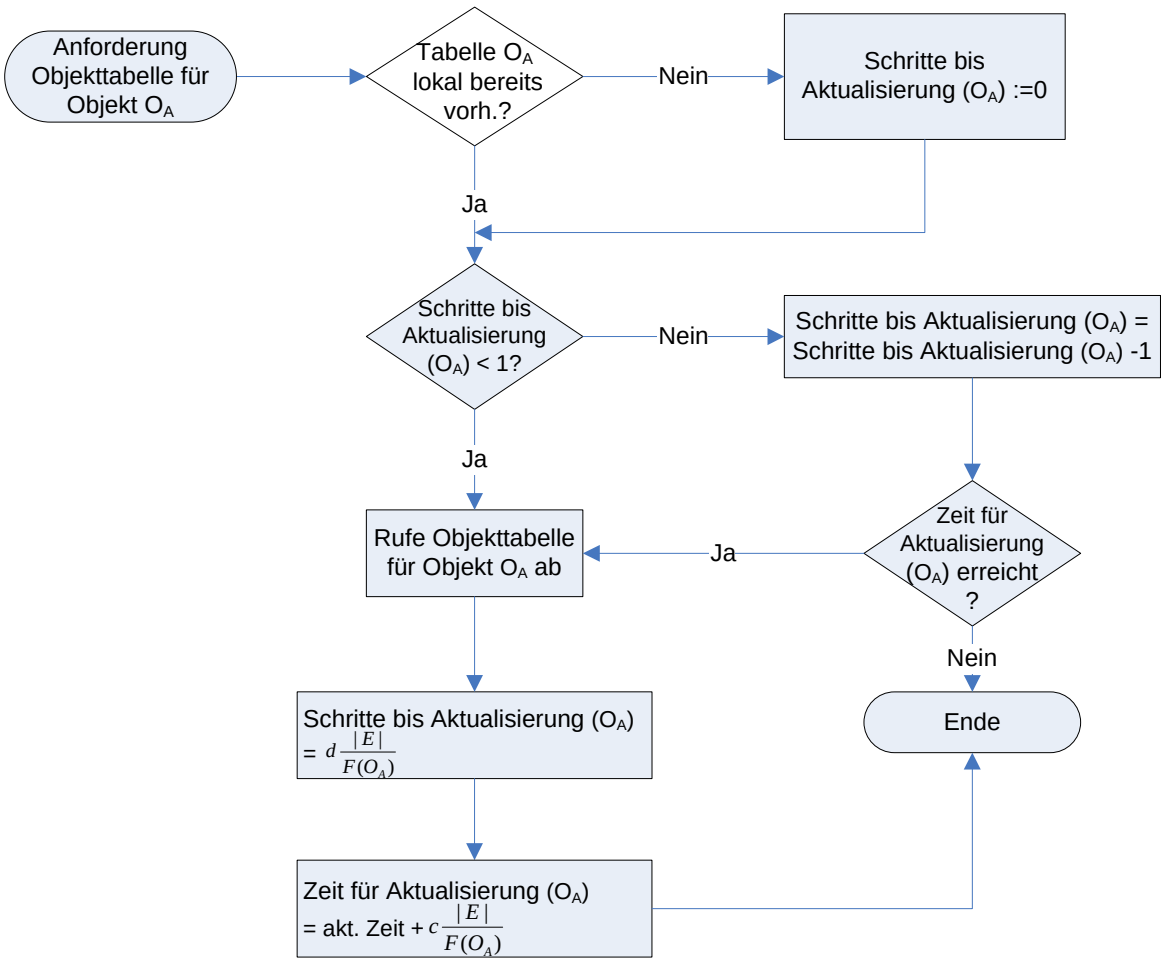

Abbildung 5.6: Aggregierte Aktualisierung von Objekttabellen

zerprofil enthaltenen Objekte müssen die neu hinzugekommenen Objekte hinzugefügt werden. Hier ist eine Priorisierung nach Relevanz des jeweils zu aktualisierenden älteren Objekts möglich; dabei ist jedoch darauf zu achten, dass hierdurch die Zusammenfassung mehrerer Aktualisierungen in einer Nachricht erschwert wird. Aus diesem Grund wird für die Implementierung im Rahmen dieser Arbeit auf eine solche Priorisierung verzichtet.

Zusammenfassend wird eine Aktualisierungsnachricht dann versendet, wenn eines der beiden folgenden Kriterien erfüllt ist:

$$\Delta t = f \cdot \log |R| \tag{5.15}$$

$$\Delta B = g \cdot \log |R| \tag{5.16}$$

mit $|R|$ Gesamtzahl versandter Aktualisierungsnachrichten, B Anzahl durch den Nutzer abgegebener Bewertungen, f und g Konstanten. Ob die Kriterien erfüllt sind, wird auch hier erst überprüft, sobald neue Objekte benutzt und bewertet wurden. Das Vorgehen beim Einfügen einer neuen Bewertung ist in Abbildung 5.7 dargestellt: Jede Bewertung, die der Nutzer abgibt, wird sofort dem lokalen Nutzerprofil hinzugefügt. Sobald im lokalen Nutzerprofil eine Bewertung länger als die Zeitspanne Δt vorliegt und noch nicht veröffentlicht ist oder mehr als ΔB (gemäß Gleichung 5.15 bzw. 5.16) Bewertungen nicht veröffentlicht sind, wird eine Aktualisierungsnachricht versendet.

Im Zusammenhang mit dem Multicast-Verfahren aus Abschnitt 5.8.3.2 werden beim *Einfügen* von Bewertungen nur zwei Nachrichten versendet. Zunächst wird unter den Schlüsseln der bereits vorher bewerteten Objekte ein Dokument mit folgendem Inhalt eingefügt:

- Nutzer-ID des bewertenden Nutzers
- Für jedes neu bewertete Objekt:
  - Identifikator des Objekts
  - Zeitpunkt der Benutzung
  - Bewertungszahl
  - Signatur über Nutzer-ID, Identifikator des neu bewerteten Objekts, Zeitpunkt der Benutzung und Bewertungszahl

Unter den Schlüsseln der neu bewerteten Objekte wird ein Dokument mit dem folgenden Inhalt eingefügt:

- Nutzer-ID des bewertenden Nutzers
- Für jedes neu bewertete und jedes vorher bewertete Objekt:

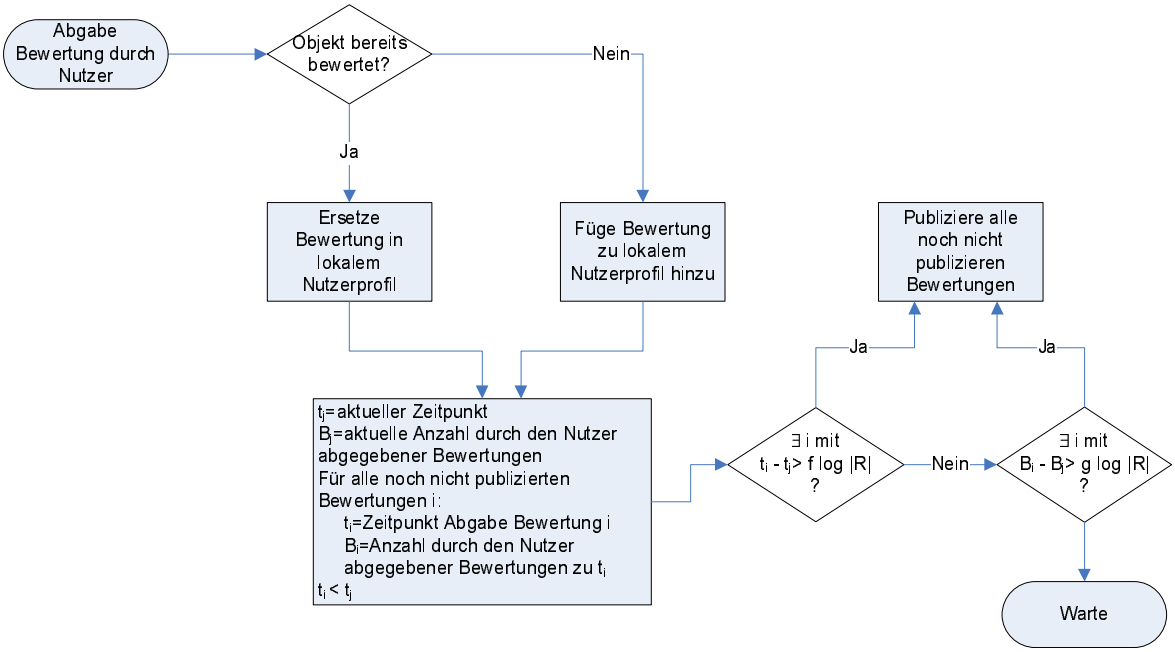

Abbildung 5.7: Einfügen neuer Bewertungen

- Identifikator des Objekts
- Zeitpunkt der Benutzung
- Bewertungszahl
- Signatur über Nutzer-ID, Identifikator des Objekts, Zeitpunkt der Benutzung und Bewertungszahl

Beim *Generieren* einer Empfehlung bleibt es bei dem in Abbildung 5.6 dargestellten Verfahren. Statt des Schritts „Rufe Objekttabelle für Objekt o ab“ wird dieses Objekt lediglich in eine Liste zu aktualisierende Objekttabellen eingetragen; erst, wenn die Prüfung aus Abbildung 5.6 für alle Objekte abgeschlossen ist, die zur Erzeugung einer Empfehlung benötigt werden, wird die Aktualisierungsanfrage für alle Objekte versandt, also die Operation *suchen(Schlüssel der eingetragenen Objekte)* ausgeführt.

5.8.3.4 Aggregierte Speicherung von Bewertungen

Der Kommunikationsaufwand lässt sich zusätzlich verringern, indem Informationen über mehrere Objekte zusammengefasst auf jeweils einem Knoten gespeichert werden. Dies lässt sich einerseits – jedoch auf Kosten der Lastverteilung – dadurch erreichen, dass die Anzahl an Knoten, die grundsätzlich Informationen speichern können, reduziert wird. Dazu könnte beispielsweise eine modifizierte Chord-ID zu jedem Schlüssel gebildet werden, indem die ursprüngliche Chord-ID ganzzahlig durch eine Konstante k geteilt wird ($\textit{Chord-ID}_{neu} = \textit{Chord-ID mod } k$).

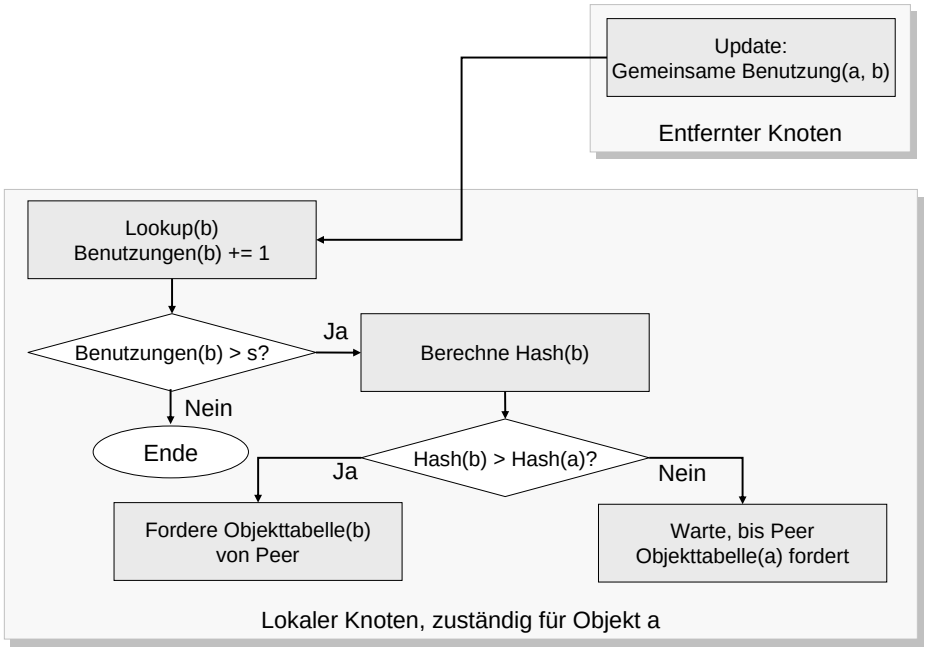

Abbildung 5.8: Aggregation von Objekttabellen

Eleganter erscheint jedoch die Zusammenfassung von Objekten, die ohnehin häufig gemeinsam abgerufen werden. Dies bietet sich beim verwendeten Ansatz schon deshalb an, weil die benötigte Information – die Häufigkeit gemeinsamer Benutzungen – ohnehin bereits vorhanden ist. Abbildung 5.8 fasst den Algorithmus zusammen, mit dem diese Aggregation durchgeführt wird. Sobald ein vorgegebener Schwellwert gemeinsamer Nutzungen überschritten wird, sendet einer der beteiligten Knoten seine Objekttabelle an den Knoten, der für das zweite Objekt zuständig ist – er verwendet dazu den Aufruf *einfügen(Schlüssel des zweiten Objekts, Objekttabelle)*. Welcher Knoten zuständig wird, wird durch den Vergleich der Hashwerte der beiden Objekt-Identifikatoren zufällig, aber nachvollziehbar bestimmt.

Der Knoten, der seine Zuständigkeit verloren hat, beantwortet künftige Anfragen mit einem Verweis auf den neuen Schlüssel. Dies erhöht zwar den Aufwand bei einer erstmaligen Anfrage, doch durch Caching der Verweisinformation können zukünftige Anfragen direkt an den neu zuständigen Knoten gerichtet werden. Jeder Knoten, der einen Verweis auf einen neuen Schlüssel erhält, speichert diesen lokal ab.

5.8.4 Angriffsvektoren und Lösungswege

In der vorgestellten Version lässt sich zwar die Funktionalität des Empfehlungssystems gewährleisten, doch sind noch keinerlei Sicherheitsmaßnahmen vorgesehen. Eine Vielzahl von

Angriffen ist denkbar. An dieser Stelle werden Abwehrmaßnahmen gegen einzelne Angriffe erörtert; eine ausführliche Analyse der Sicherheit des Gesamtsystems folgt in Kapitel 6.

Angriffe durch speichernde Knoten Die erste Kategorie von Angriffen ist nur den Knoten möglich, die für die Speicherung einer Objekttabelle zuständig sind. Diese können:

- Eine höhere Anzahl gemeinsamer Objektbenutzungen mit bestimmten anderen Objekten vortäuschen, als diese tatsächlich vorliegt.
- Eine niedrigere Anzahl gemeinsamer Objektbenutzungen mit bestimmten anderen Objekten vortäuschen, als diese tatsächlich vorliegt.
- Eine höhere Gesamtzahl an Benutzungen des von ihnen verwalteten Objekts vortäuschen, als sie tatsächlich vorliegt.
- Eine niedrigere Gesamtzahl an Benutzungen des von ihnen verwalteten Objekts vortäuschen, als sie tatsächlich vorliegt.
- Einen Teil der empfangenen Anfragen – oder alle – ignorieren.

Angriffe durch andere Knoten Auch andere Knoten können das System erheblich beeinträchtigen. Im Wesentlichen kann ein Angreifer gemeinsame Objektbenutzungen veröffentlichen, die nicht stattgefunden haben – oder in höherer Anzahl, als sie stattgefunden haben.

Im Folgenden werden all diese Angriffsmöglichkeiten und die angewandten Lösungswege diskutiert.

5.8.4.1 Angriffe durch speichernde Knoten

In einem ersten Schritt soll der Anreiz eines Knotens reduziert werden, Empfehlungen für ein Objekt zu beeinflussen. Wer ein solches Interesse hat, wird im Allgemeinen daran interessiert sein, Empfehlungen für ein *bestimmtes* Objekt zu erreichen oder zu verhindern. Ziel muss also sein, die Auswahl der Chord-ID, für die ein Knoten zuständig sein soll, durch diesen Knoten zu erschweren. Wie dies möglich ist, wird bereits in Abschnitt 2.1.7 diskutiert.

Zusätzlich sollen jedoch Gegenmaßnahmen gegen einzelne Angriffe, wie sie oben aufgeführt wurden, ergriffen werden.

Übertreiben der Anzahl gemeinsamer Objektbenutzungen Ein Angreifer, der ein Objekt bewerben will, kann dies dadurch tun, dass er auf Anfrage behaupten, dieses sei sehr häufig mit anderen Objekten benutzt worden. Es bringt dem Angreifer dabei jedoch keinen Vorteil, die Objekttabelle dieses Objekts zu verwalten: Wer die entsprechende Objekttabelle abruft,

tut dies, weil das Objekt bereits Teil seines Profils ist und er nach Empfehlungen für *andere* Objekte sucht. Doch auch, wenn der Angreifer andere Objekttabellen verwaltet, kann dem Angriff vorgebeugt werden: Zu diesem Zweck muss der Knoten, der eine Objekttabelle speichert, nicht nur die Anzahl gemeinsamer Objektbenutzungen speichern, sondern auch die signierten Aktualisierungsnachrichten der Knoten, die das Objekt verwendet haben. Auf diese Weise kann er beweisen, dass nicht weniger gemeinsame Objektbenutzungen vorliegen, als von ihm behauptet.

Untertreiben der Anzahl gemeinsamer Objektbenutzungen Der umgekehrte Angriff – das Zurückliefern einer kleineren Anzahl gemeinsamer Objektbenutzungen, als sie tatsächlich vorliegen – ist schwieriger aufzudecken. Ein Angreifer könnte daran Interesse haben, um eine Empfehlung für ein anderes Objekt zu provozieren. Zur Überprüfung kann man sich jedoch die Symmetrie der gespeicherten Informationen zunutze machen: Wurde Objekt A 10mal gemeinsam mit Objekt B genutzt, so gilt dies natürlich auch umgekehrt. Ein überprüfender Knoten kann also den für Objekt B verantwortlichen Knoten befragen und sich von diesem auch belegen lassen, dass der durch ihn gelieferte Wert nicht zu hoch ist.

Um diese Überprüfung effizient zu gestalten, wird sie jedoch nicht für alle Objekte durchgeführt. Lediglich Objekte, die dem Nutzer tatsächlich vorgeschlagen werden, werden auf diese Weise überprüft.

Übertreiben der Gesamtzahl an Objektbenutzungen Die Information, wie oft ein Objekt insgesamt benutzt wurde, fließt in die Empfehlungsberechnung ebenfalls mit ein. Der Knoten, der die jeweilige Objekttabelle speichert, kann analog zu den einzelnen Einträgen der Tabelle auch die Gesamtsumme an Objektbenutzungen beweisen, indem er alle signierten Aktualisierungsnachrichten, aus denen sich die Zahl ergibt, einem überprüfenden Knoten zur Verifikation sendet. Da es sich jedoch um eine sehr große Zahl an Dokumenten handeln kann, sollte eine solche Überprüfung nicht bei jeder Anfrage an einen Knoten durchgeführt werden.

Untertreiben der Gesamtzahl an Objektbenutzungen Schwieriger stellt sich das Problem dar, dass ein Knoten die Gesamtzahl an Objektbenutzungen auch untertreiben kann. Werden beispielsweise einzelne Zeilen der Objekttabelle als leer zurückgemeldet, so kann gerade nicht mehr durch Anfrage bei einem anderen Knoten überprüft werden, ob diese Aussage korrekt ist. Eine Lösung kann hier nur in redundanter Speicherung von Daten liegen. Auch hier soll aber auf den Abgleich bei jeder Anfrage verzichtet werden, sondern nur eine stichprobenartige Überprüfung vorgenommen werden.

Ignorieren von Anfragen Ignoriert ein Knoten Anfragen, so kann der anfragende Knoten dies zwar an ausbleibenden Bestätigungen bemerken – sicher verhindern lässt sich ein sol-

ches Verhalten jedoch nicht. Auch hier wird die Funktionalität des Systems also durch die redundante Speicherung von Daten sichergestellt.

5.8.4.2 Sonstige Angriffe

Problematischer als die Angriffe der für ein Objekt jeweils verantwortlichen speichernden Knoten sind solche, die allen Systemteilnehmern möglich sind: Diese betreffen nicht mehr nur je ein Objekt, sondern potentiell alle dem System bekannten.

Einfügen inkorrekter Objektbenutzungsangaben Das Einfügen inkorrekter Objektbenutzungsangaben kann nicht direkt sanktioniert werden – welche Objekte ein Systemteilnehmer benutzt hat, ist durch das System nicht nachvollziehbar. Denkbar sind lediglich Plausibilitätsprüfungen (die Anzahl an Objekten, die ein Nutzer in der Praxis benutzen kann, ist begrenzt) und Bewertungen aufgrund der Qualität der Empfehlungen, die sich bei Berücksichtigung des entsprechenden Nutzers ergibt. Hier besteht allerdings das Problem, dass Nutzern mit ungewöhnlichem Bewertungsprofil der Anreiz genommen werden könnte, dieses wahrheitsgemäß zu offenbaren.

Der Anreiz zum Einfügen inkorrekter Angaben könnte jedoch – wenn der Sybil-Angriff, wie in Abschnitt 2.1.7 diskutiert, vermieden wird – auch dadurch reduziert werden, dass Bewertungen eines Nutzers geringer gewichtet werden, wenn dieser viele Bewertungen abgegeben hat. Einfach zu realisieren ist dies bei Verzicht auf die Anonymitätsunterstützung. So kann für jeden Knoten A ein weiterer Knoten B bestimmt werden, der für die Speicherung der Anzahl bereits vorgenommener Bewertungen zuständig ist. In diesem Fall legt A den Hash-Wert der aktualisierten Objekttabelle it zur Signatur bei B vor; B signiert sodann die Zeichenkette ($A \mid hash(it) \mid$ Zeitstempel $\mid$ Anzahl von A benutzter Objekte). Die so signierte Zeichenkette wird bei den für die Speicherung zuständigen Knoten vorgelegt, die die Signatur von A prüfen und, wenn der enthaltene Zeitstempel aktuell ist, ihre Objekttabellen aktualisieren. Aus Sicherheitsgründen bietet es sich an, statt eines einzelnen Knotens B mehrere einzusetzen; ein nicht kooperativer Knoten könnte sonst leicht durch Verweigern der Signatur die Teilnahme des korrespondierenden A-Knotens am System verhindern.

Konkret wird die geringere Gewichtung von Knoten, die sehr viele Bewertungen abgegeben haben, realisiert, indem das in Gleichung 5.6 gegebene Relevanzmaß modifiziert wird. Die Funktion $F(O_A)$, die in der ursprünglichen Version lediglich die Anzahl der Benutzungen des Objekts A angibt, wird dazu umdefiniert. Statt die Anzahl der Nutzerprofile $L_i(O_A)$ anzugeben, in denen das Objekt vorkommt, werden die Nutzer mit einem Gewicht versehen:

$$F(O_A) = \sum_{i=1}^{N} w_i(L_i) \cdot L_i(O_A), \tag{5.17}$$

wobei $w_i(L_i)$ eine Gewichtungsfunktion ist, die dazu führt, dass Nutzerprofile mit vielen Einträgen geringer gewichtet werden als solche mit wenigen Einträgen.

Als einfache Funktion wird dabei vorgeschlagen

$$w_i(L_i) = \frac{1}{1 + \log |L_i|} \tag{5.18}$$

Diese Gewichtungsfunktion wird auch in der Evaluierung des vorgestellten Ansatzes zugrunde gelegt.

5.8.4.3 Vertrauen als Empfehlungsgrundlage

Zusätzlich kann bei der Empfehlungsberechnung das Vertrauen in andere Knoten mit einfließen. Der einfachste Weg, dies zu erreichen, ist eine Modifikation der Gewichtungsfunktion aus Gleichung 5.18 durch Gewichtung mit dem Vertrauen t_i in den jeweiligen Knoten:

$$w_i(L_i) = t_i \cdot \frac{1}{\log |L_i|} \tag{5.19}$$

Der Vertrauenswert kann sich dabei wiederum aus der Verwendung eines Reputationssystems ergeben.

Ähnlich dem in Abschnitt 5.9.1 Verfahren kann auch hier Reputation in Abhängigkeit von der Qualität einer errechneten Empfehlung generiert werden. Wiederum wird die Empfehlung sowohl ohne als auch mit dem Beitrag des entsprechenden Nutzers berechnet; ergibt sich durch die Berücksichtigung des Nutzers bei der Empfehlungsberechnung eine Verbesserung, wird eine positive Bewertung abgegeben, sonst eine negative.

5.8.4.4 Reputation und Datenschutz

Reputationssysteme können auch aus einem anderen Gesichtspunkt heraus hilfreich sein. Wie sich gezeigt hat, ist eine Überprüfung der Angaben von Knoten in einigen Fällen zwar möglich, soll aufgrund des hohen Aufwands aber nicht bei jedem Abruf, sondern nur stichprobenartig vorgenommen werden. Es stellt sich somit jedoch die Frage, wie auf Angriffe reagiert werden soll. Eine natürliche Reaktion auf den Angriff eines Knotens ist es, diesem Knoten nicht mehr zu vertrauen – um dies zu kommunizieren, ist der Einsatz eines Reputationssystems unerlässlich.

Wie bereits erwähnt, soll an dieser Stelle kein neuer Reputationsmechanismus entwickelt werden – vielmehr stellt sich die Frage, wie ein bestehendes Reputationssystem einbezogen werden kann, ohne dabei die hohen Anforderungen an den Datenschutz zu verletzen. Die Grundidee besteht dabei darin, einem Knoten zwar eine persistente Identität zuzuweisen und diese auch für das Reputationssystem zu verwenden, jedoch in den Objekttabellen auf

die Verwendung dieser Identitäten zu verzichten. Aufgrund des damit verbundenen Kommunikationsaufwands bleibt auch dieses Protokoll optional.

Schutz der Identität bei der Aktualisierung von Objekttabellen Der Prozess beim Aktualisieren einer Objekttabelle verläuft wie folgt:

Die Aktualisierungsnachricht besteht aus zwei Teilen: Dem bestehenden Nutzerprofil, aus dem sich ergibt, an welche Knoten die Nachricht geschickt wird, und den l neuen Objekten, die der Nutzer hinzufügen möchte.

Beim Versand einer Aktualisierungsnachricht führt der jeweilige Knoten N nun den folgenden Algorithmus durch:

```
Für jedes neu bewertete Objekt k
   Für jedes vorher bewertete Objekt p
       berechne hash(k, p, timestamp, salt),
```

mit $salt$ Zufallszahl. Alle berechneten Hash-Werte für ein festes k werden an den Knoten H^k gesendet, der für den folgenden Schlüssel verantwortlich ist: (Erste $\frac{t}{2}$ bits von hash(Identifikator des neu bewerteten Objekts k) | Erste $\frac{t}{2}$ bits von hash(N)), mit '|' Konkatenationsoperator, t Länge einer Knoten-ID in bits.

H^k überprüft sodann, ob er für Anfragen von N zuständig ist, indem er ebenfalls die ersten $\frac{t}{2}$ bits von hash(N) berechnet. Ist das der Fall, prüft er, ob er schon einmal eine Anfrage von N erhalten hat – ist das der Fall, verwirft er die Anfrage. Andernfalls signiert H^k die Zeichenketten (H^k | $hash(k, p, timestamp, salt)$ | timestamp | lokal eindeutiger Bezeichner für N | Anzahl vorher von N bewerteter Objekte | Reputation von N). Er speichert die Host-ID des Knotens N und den lokal eindeutigen Bezeichner für N in einer Datenbank und sendet die signierten Zeichenketten an N. Der Knoten N sendet daraufhin die Aktualisierungsnachrichten; jede enthält ein neu bewertetes Objekt und einen Teil der vorher bewerteten Objekte. Für jeden Eintrag werden der Salt-Wert und die von H^k empfangenen Informationen, einschließlich des bei H^k lokal eindeutigen Bezeichners für N, mitgeschickt. Für den ersten Overlay-Hop jeder Aktualisierungsnachricht wird Onion Routing als Anonymisierungsprotokoll verwendet, sodass für den empfangenden Knoten nicht ersichtlich ist, wer die Nachricht abgesendet hat. Der für die jeweilige Objekttabelle zuständige Knoten wiederum prüft, ob der Knoten H^k tatsächlich für das Objekt zuständig gewesen sein kann. Falls dies der Fall ist, wird die Signatur geprüft und der Eintrag aus benutztem Objekt und den von H^k versandten Informationen in der Objekttabelle gespeichert.

Mit Hilfe dieses Protokolls werden zweierlei Ziele erreicht: Zum einen wird verhindert, dass ein Knoten mehrfach die gleiche Objektkombination angibt. Zum anderen wird die Reputation eines Knotens bei der jeweiligen Objektkombination vermerkt. Beides ist möglich, ohne die Identität dieses Knotens zu enthüllen. Es bleibt jedoch die Frage, wie diese Reputation bei Bedarf aktualisiert werden kann.

Aktualisierung von Reputationsinformation Offensichtlich kann es vorkommen, dass die Reputation eines Knotens aktualisiert werden muss, nachdem dieser eine Bewertung abgegeben hat. Jedoch kennt ein Knoten KB, der auf eine Bewertung zugreift, nicht die Identität des Bewertenden. Um basierend auf einer Transaktion sein Vertrauen zum Ausdruck zu bringen, sendet er deshalb eine Nachricht an den Knoten H^k, der die Bewertung – nach Abwarten eines zufälligen Zeitraums Δt – an das Reputationssystem weiterreicht. Folge dieses Verfahrens ist jedoch, dass das Reputationssystem die Bewertung behandeln muss, als habe der Knoten H^k sie aufgrund eigener Erfahrungen abgegeben. Wird stattdessen die Identität von KB verwendet, können Dritte eine Zuordnung zwischen dieser und der Identität des Bewerteten herstellen.

Da die Identität eines Knotens jedoch für jedes Objekt von einem anderen Knoten verwaltet wird, kann im Fall eines Knotens, der unzutreffende Objektbewertungen abgibt, das Reputationssystem dennoch effektiv eingesetzt werden.

5.9 Erweiterungen

Neben den beschriebenen Mechanismen sind Erweiterungen denkbar, die für die grundlegende Funktionalität der vorgestellten Verfahren nicht notwendig sind, aber Verbesserungen in verschiedenen Eigenschaften bewirken können.

5.9.1 Vertrauen, Sicherheit und Datenschutz im nutzerbasierten Ansatz

Der nutzerbasierte Ansatz lässt sich durch das Einbeziehen von Vertrauensbeziehungen erweitern. Erweisen sich die Empfehlungen eines Knotens als hilfreich, so ist dies ein Anlass, diesem Knoten zukünftig mehr Vertrauen entgegenzubringen und sie folglich mit einem höheren Gewicht in die eigene Entscheidungsfindung einfließen zu lassen.

Wiederum kann dieses Vertrauen zunächst lokal gespeichert werden; der Einsatz eines Reputationssystems, um lokale Vertrauensentscheidungen auch anderen Netzteilnehmern kommunizieren zu können, ist aber eine wünschenswerte Ergänzung. Das Reputationssystem selbst ist nicht Fokus dieser Arbeit. In der Literatur ist bereits eine Reihe von Reputationssystemen für Peer-to-Peer-Systeme vorgestellt worden – genannt seien hier insbesondere das Eigentrust-System [KaSGM03] sowie [KiPe03], [CDVP+02] und [GuJA03] sowie (auf Empfehlungssysteme bezogen) [Zieg05].

Die Verwendung eines Vertrauenswerts in einem Empfehlungsalgorithmus ist dabei zunächst unproblematisch; eine einfache Gewichtung von Eingabegrößen mit dem Vertrauenswert des jeweiligen Nutzers reicht in der Regel aus. Fraglich ist jedoch, wie ein Vertrauenswert konkret berechnet werden kann. Verschiedene Verfahren sind denkbar:

- Vertrauen kann durch persönliche Kontakte zwischen den Teilnehmern des Systems entstehen. Zusätzliche Vertrauenswerte können durch manuelle Bewertungen seitens

der Nutzer generiert werden, die beispielsweise aufgrund ähnlicher Bewertungsprofile entscheiden. Dieser Weg eignet sich offensichtlich nur für eine geringe Anzahl von Kommunikationspartnern.

- Vertrauenswerte können auch auf Basis erzeugter Empfehlungen vergeben werden: Führt die Berücksichtigung des Profils eines Knotens bei der Empfehlungsberechnung zu einer Einschätzung, die von der danach durch den Nutzer des empfohlenen Guts vorgenommenen Einschätzung abweicht, so führt dies zu einer Vertrauensreduktion, andernfalls wird der Vertrauenswert erhöht. Hier besteht die Gefahr, dass Nutzern das Vertrauen lediglich aufgrund ungewöhnlicher Vorlieben entzogen wird. Dennoch wird der Ansatz in der folgenden Art und Weise verfolgt: Die Bewertungen von Objekten werden wie oben beschrieben berechnet. Diese Berechnung wird im Anschluss pro taste buddy einmal wiederholt, wobei das Profil des jeweiligen taste buddy nicht in die Berechnung einfließt. Nun wird der Betrag d der Differenz aus der durch den Nutzer selbst vorgenommenen Bewertung $B^*_{O_A(i)}$ und der ursprünglichen Bewertung $B_{O_A(i)}$ des Empfehlungssystems gebildet. Zusätzlich berechnet das System den Betrag $d^{\setminus j}$ der Differenz zwischen $B^*_{O_A(i)}$ und der Bewertung ohne taste buddy j $B_{O_A^{\setminus j}(i)}$. Ist nun $d - d^{\setminus j} > 0$, so hat sich die Empfehlungsqualität durch Einbeziehung von taste buddy j verbessert, und der lokale Vertrauenswert wird erhöht; sonst wird er gesenkt.

Die auf diese Art gewonnenen Vertrauenswerte können verwendet werden, um mit Hilfe des in Abschnitt 4.4.2 beschriebenen Systems die Kenntnisnahme von Bewertungen durch andere Knoten einzuschränken.

5.9.2 Kategoriebildung beim objektbasierten Ansatz

Die Suche nach einer Empfehlung wird in aller Regel nicht dem Ziel dienen, Empfehlungen zu *irgendeinem* Objekt zu erhalten; vielmehr hat der Suchende Interesse an einer bestimmten Kategorie von Objekten. In der Literatur bislang betrachtete Peer-to-Peer-Empfehlungssysteme betrachten nur Empfehlungen innerhalb einer bestimmten Domäne, doch bedeutet dies eine erhebliche Einschränkung der Anwendbarkeit solcher Systeme.

Im vorliegenden System soll dementsprechend eine Empfehlungsberechnung in Abhängigkeit von Objektkategorien ermöglicht werden. Dabei stellen sich zwei wesentliche Herausforderungen:

- Die Bildung von Objektkategorien muss ermöglicht werden.
- Die Empfehlungserzeugung muss so modifiziert werden, dass sie die gewünschte Zielkategorie berücksichtigt.

Beide Probleme werden in den folgenden Absätzen erörtert.

Bildung von Objektkategorien Objektkategorien sollen in eine Baumstruktur eingeordnet werden: Ausgehend von einer Wurzelkategorie werden dabei jeweils Unterkategorien definiert. Jede Kategorie gehört zu genau einer Oberkategorie, kann jedoch beliebig viele Unterkategorien haben. Der Baum kann im System im einfachen Fall statisch vorgegeben werden. Dazu ist es möglich, bestehende Taxonomien von Objekten heranzuziehen. Eine solche Taxonomie wird beispielsweise in [LAKV+01] herangezogen.

Alternativ oder ergänzend besteht die Möglichkeit, eine Taxonomie durch die Nutzer des Systems pflegen zu lassen, wie dies beispielsweise im Wikipedia-Projekt (`www.wikipedia.org`) der Fall ist. Das Hinzufügen und Entfernen von Kategorien könnte dabei durch einen Abstimmungsprozess ermöglicht werden. Zur Speicherung der Taxonomie ist ein System denkbar, das sich am Domain Name System [RFC1034] orientiert. Die konkrete Implementierung eines solchen Systems ist jedoch nicht Gegenstand der vorliegenden Arbeit.

Empfehlungserzeugung Bei der Empfehlungserzeugung gibt es grundsätzlich zwei Möglichkeiten. So können einerseits Objekttabellen aller Kategorien berücksichtigt werden; andererseits kann auch auf die Aktualisierung von Objekttabellen nicht gewünschter Kategorien verzichtet werden. Wie sich beides auf die Empfehlungsqualität auswirkt, ist in der Literatur unzureichend untersucht. Um den Kommunikationsaufwand gering zu halten, sollte die Suche jedoch zumindest auf einen Teilbaum des oben erwähnten Kategorienbaums beschränkt werden.

5.10 Fazit

In diesem Kapitel wurden zwei Anwendungsfälle von Empfehlungssystemen identifiziert. Für den Anwendungsfall der verteilten Speicherung von Bewertungen wurde ein System entworfen, das auf dem strukturierten Overlay-Netz Chord aufbaut. Die Empfehlung dem Nutzer unbekannter Objekte kann auf zwei verschiedenen Wegen realisiert werden. Ein Ansatz eignet sich für unstrukturierte wie auch für strukturierte Netze, stellt jedoch nur einen Ausschnitt potentiell relevanter Daten zur Verfügung. Durch Ausnutzung der Eigenschaften strukturierter Netze ermöglicht der zweite Ansatz eine vollständigere Datengrundlage für die Empfehlungsberechnung.

Beide Ansätze sollen durch diverse Protokollerweiterungen in ihren Eigenschaften bezüglich Sicherheit, Datenschutz und Skalierbarkeit verbessert werden. Ein hybrider Ansatz kann zudem die Vorteile beider Ansätze in Teilen vereinen. Die Ergebnisse des vorliegenden Kapitels sind in Teilen in [Sorg07a] veröffentlicht.

Im folgenden Kapitel wird analysiert, inwieweit diese Verbesserungsvorschläge die identifizierten Anforderungen erfüllen.

Kapitel 6

Evaluation

In diesem Kapitel wird der Entwurf, wie er im vorherigen Kapitel dargestellt wurde, anhand verschiedener Kriterien bewertet. Zunächst wird eine prototypische Implementierung dargestellt, die zeigt, dass das entworfene Empfehlungssystem grundsätzlich in die Praxis umgesetzt werden kann. Die Bewertung aus technischer Sicht wird im Anschluss nach drei Kriterien vorgenommen, die in den folgenden Abschnitten betrachtet werden:

- Zum einen wird die Qualität der errechneten Empfehlungen der Anwendungen des Kollaborativen Filterns betrachtet. Kann nur ein Teil der Eingabedaten zur Verfügung gestellt werden, die in einem zentralen System verfügbar wären, ist mit einer Verschlechterung der Qualität zu rechnen, die es aber in Grenzen zu halten gilt. Die Qualität wird simulativ ermittelt, da so reproduzierbare Ergebnisse erlangt werden können.
- Den zweiten Schritt bilden Betrachtungen zur Skalierbarkeit. Auch hier bildet eine Simulation des Systems die Grundlage; die Überprüfung der Skalierbarkeit mittels einer prototypischen Implementierung würde sich im Vergleich deutlich aufwendiger gestalten.
- Den Abschluss bilden Sicherheitsbetrachtungen. Diese basieren auf theoretischen Überlegungen.

Da ein Großteil der technischen Evaluation auf Simulationen des Systems aufbaut, soll zunächst deren Aufbau beschrieben werden. Im Anschluss werden Bewertungsmetriken vorgestellt, die der Einordnung der erreichten Empfehlungsqualität dienen können. Schließlich werden die einzelnen oben genannten Bewertungskriterien betrachtet.

6.1 Prototypische Implementierung

Die prototypische Implementierung eines Empfehlungssystems dient dazu, die praktische Anwendbarkeit des entworfenen Systems zu überprüfen. Auf Grundlage des SESAM-Basissystems [CDHS+05] setzt sie die verteilte Speicherung von Bewertungsdokumenten

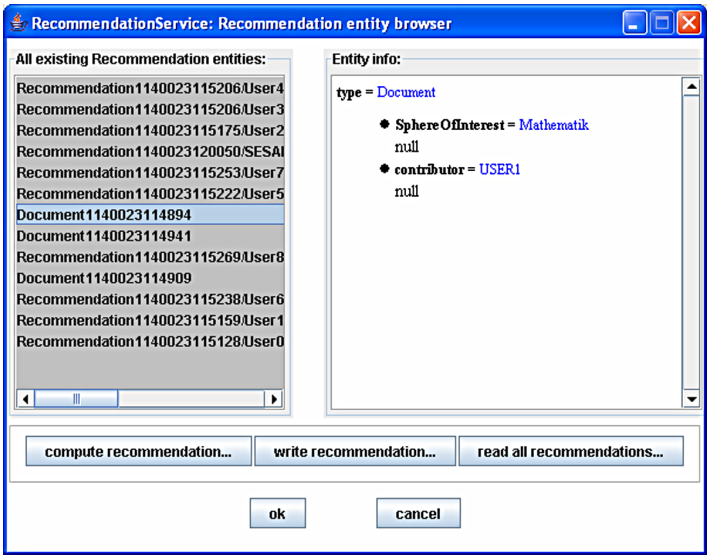

Abbildung 6.1: Prototypische Implementierung

sowie die Grundversion objektorientierten Kollaborativen Filterns um – jedoch ohne die funktionalen Erweiterungen, die in der Simulation evaluiert werden. Wesentliche Teile der Implementierung gehen dabei auf eine Studienarbeit [Scan06] zurück, die im Rahmen der vorliegenden Arbeit entstand.

Das SESAM-Basissystem stellt bereits ein Äquivalent zur Datenspeicherungsschicht des in Kapitel 5 beschriebenen Entwurfs zur Verfügung: Der *Dokumentendienst* ermöglicht es, Dokumente – so auch Bewertungsdokumente – zu speichern, zu löschen, und abzurufen. Die funktionellen Anforderungen an die Datenspeicherungsschicht werden durch den Dienst also erfüllt. Er wird daher anstelle der beschriebenen Datenspeicherungsschicht herangezogen. Inwieweit die diskutierten Datenschutzanforderungen erfüllt werden, hängt von der konkreten Implementierung der darunter liegenden Schichten ab. Um die Durchführbarkeit des gewählten Ansatzes zu demonstrieren, reicht das SESAM-Basissystem jedoch jedenfalls aus.

Der Empfehlungsdienst beinhaltet die Komponente *RecommendationCache*, die lokale Kopien von Bewertungsdokumenten vorhält, sowie den *RecommendationGenerator*, der den eigentlichen Empfehlungsalgorithmus beinhaltet. Dieser Algorithmus ist dabei austauschbar und kann zur Laufzeit ausgewählt werden. Die Nutzeroberfläche – in einem Beispiel in Abbildung 6.1 dargestellt – erlaubt es Nutzern, Bewertungen abzugeben, nach Bewertungsdokumenten zu suchen und Empfehlungen zu berechnen.

Als Besonderheit der Implementierung ist die Möglichkeit des Einsatzes einer Taxonomie hervorzuheben. Im SESAM-Basissystem werden Ontologien eingesetzt, also (formale) Spezifikationen einer Begriffsbildung [Grub93]. Wesentlicher Mechanismus dieser Spezifikation ist die Einordnung in eine Taxonomie. Die Verwendung von Ontologien wurde für

die verwendeten Bewertungsdokumente auch in das prototypisch implementierte Empfehlungssystem übernommen: Dazu wurde eine einfache Beispieltaxonomie für die Ansicht zu empfehlender Objekte eingeführt. Auf diesem Weg lässt sich die Suche nach Dokumenten einschränken, so dass nur Bewertungen bestimmter Kategorien von Objekten berücksichtigt werden. Beispielsweise können so die Bewertungen aller Sportartikel gesucht werden.

6.2 Simulationen

6.2.1 Verwendeter Datensatz

Eine Reihe verschiedener Datensätze ist in der Literatur für die Evaluation von Empfehlungssystemen herangezogen worden. Die gängigsten sind der EachMovie-Datensatz, der Jester-Datensatz und der Movielens-Datensatz [HKTR04]. In [BEKR06] sind diese Datensätze analysiert und gegenübergestellt:

- Der EachMovie-Datensatz besteht aus insgesamt 2.711.718 Bewertungen von 74.424 Nutzern über 1.649 Objekte (Filme). Dieser Datensatz ist jedoch nicht mehr öffentlich verfügbar.
- Der Jester-Datensatz besteht aus 3.519.449 Bewertungen von 48.483 Nutzern über 100 Objekte (Witze).
- Der MovieLens-Datensatz steht in verschiedenen Varianten zur Verfügung; es handelt sich hierbei um verschieden große Ausschnitte einer Datenbasis von Filmbewertungen. Die umfangreichste davon („Movielens Million“) enthält 1.000.209 explizite Bewertungen auf einer ganzzahligen Skala von 1–5, die 6.040 Nutzer über 3.952 Filme abgegeben haben.

Außer den in [BEKR06] dargestellten Datensätzen ist noch der Bookcrossing-Datensatz [ZMKL05] zu erwähnen. Dieser enthält 1.149.780 explizite Bewertungen von 278.858 Nutzern über 271.379 Objekte (Bücher).

Für die Simulation der untersuchten Ansätze wurde der durch das GroupLens-Projekt[1] zur Verfügung gestellte Movielens-Datensatzes ausgewählt. Bei diesem Datensatz ist im Gegensatz zu den anderen genannten Datensätzen jede Bewertung mit einem Zeitstempel versehen; dies ermöglicht eine Einbeziehung des zeitlichen Ablaufs der Bewertungsabgabe in die durchgeführten Simulationen. Der Zeitraum, in dem die Bewertungen gesammelt wurden, reicht von April 2000 bis Februar 2003. Demographische Informationen über die Nutzer sowie eine Einordnung der Filme in Genres sind ebenfalls enthalten, wurden jedoch nicht verwendet.

[1] `http://www.grouplens.org`, abgerufen am 20. März 2007

In [BEKR06] sind einige statistische Merkmale des Movielens-Datensatzes zusammengefasst: Im Durchschnitt hat jeder Nutzer 165,6 Bewertungen abgegeben; dies führt zu einer Bewertungsdichte (Quotient aus tatsächlich abgegebenen und theoretisch möglichen Bewertungen) von 4,19%. Die durchschnittliche Bewertung liegt bei 3,580, bei einer Standardabweichung von 0,935.

In den Simulationen wurde jedem im Movielens-Datensatz vorhandenen Nutzer ein zufällig ausgewählter Knoten im Peer-to-Peer-System zugeordnet – jedoch nicht umgekehrt, so dass stets mehr Knoten vorhanden waren als bewertende Nutzer: Zwar ist denkbar, dass in der Praxis mehrere Nutzer sich einen Knoten teilen. Dies dürfte aber nicht der Regelfall sein. Der Kommunikationsaufwand des Peer-to-Peer-Systems würde sich dadurch zudem verringern, so dass jedenfalls keine Verschlechterung der Skalierbarkeit des Systems im Vergleich zur simulierten Variante zu erwarten ist.

Die Bewertungszeitpunkte wurden gemäß der Reihenfolge gewählt, wie sie durch die Zeitstempel des MovieLens-Datensatzes vorgegeben ist. Die Simulationen wurden mit der vollen Nutzerzahl – also 6.040 Nutzern – und 10.000 Knoten durchgeführt. Für einen Teil der durchgeführten Simulationen wurde jedoch nur ein Ausschnitt der im Datensatz vorhandenen Objekte betrachtet – beispielsweise die Objekte mit den Identifikatoren 1–100. In diesem Fall wurden die dargestellten Ergebnisse als Mittelwert aus jeweils mehreren solcher Teilmengen gewonnen. Das Vorgehen erlaubte es, insgesamt eine größere Anzahl an Simulationen durchzuführen.

Ergänzend wurde für einzelne Simulationen ein weiterer Datensatz herangezogen. Hierzu kamen nur der Jester- und der Bookcrossing-Datensatz in Frage. Der Jester-Datensatz enthält allerdings nur 100 Objekte, was seine Aussagekraft zumindest für den objektbasierten Ansatz einschränkt. Daher wurde der Bookcrossing-Datensatz [ZMKL05] herangezogen. Da auch in diesem Datensatz keine Bewertungszeitpunkte enthalten sind, beschränkten sich diese Simulationen auf die Ermittlung der Empfehlungsqualität mittels der betrachteten Empfehlungsalgorithmen. Der Datensatz enthält insgesamt 1.149.780 Bewertungen von 278.858 Nutzern über 271.379 Objekte (Bücher). Zum Teil handelt es sich dabei um explizite, von den Nutzern auf einer Skala von 1–10 vorgenommene Bewertungen; es sind aber auch implizite Bewertungen enthalten, die lediglich das Interesse eines Nutzers an einem Buch ausdrücken.

6.2.2 Simulator

Die Empfehlungsqualität und der Kommunikationsaufwand, der beim Einsatz der Anwendungen für kollaboratives Filtern entsteht, wurden simulativ untersucht. Dabei wurde der in der Programmiersprache Java implementierte, ereignisbasierte Simulator *PlanetSim* [GPMP+05] eingesetzt; er wurde aufgrund seiner Skalierbarkeit – also der Möglichkeit,

ihn auch für die Simulation sehr großer Netze einzusetzen – ausgewählt[2]. Als Bestandteil des PlanetSim-Simulators ist das Chord-Protokoll implementiert. Es besteht somit die Möglichkeit, einen Chord-Ring aufzubauen, Knoten beitreten zu lassen und Nachrichten an Knoten zu senden, die für einen bestimmten Schlüssel zuständig sind. Es handelt sich um einen Overlay-Simulator; die Charakteristika des Underlays, also des IP-Netzes, das die Basis für das betrachtete Peer-to-Peer-System bildet, werden (trotz entsprechender Erweiterbarkeit des Simulators) nur unzureichend modelliert. So können insbesondere keine Latenzen bei der Kommunikation zwischen den einzelnen Knoten berücksichtigt werden; möglich ist hier lediglich eine Abschätzung aufgrund der Anzahl auf Transportschicht versandter Nachrichten. Diese Einschränkung kann jedoch für die Betrachtung einer Peer-to-Peer-Anwendung in Kauf genommen werden.

Die dargestellten Simulationen liefen jeweils wie folgt ab: Zunächst wurden 10.000 Knoten instantiiert. Diese Anzahl ergibt sich daraus, dass einerseits ein möglichst großes System untersucht werden sollte, andererseits aber die Ressourcenanforderungen der Simulation keine beliebige Steigerung zulassen.

Die Knoten traten nacheinander dem Chord-Overlay-Netz bei, wobei vollständig auf die Chord-Implementierung des PlanetSim-Simulators zurückgegriffen wurde. Den Knoten wurde dann – unabhängig von ihren Chord-IDs – Nummern von 1 bis 10.000 zugewiesen. Im verwendeten Movielens-Datensatz sind Nutzer ebenfalls durch Nummern repräsentiert. Der Datensatz wurde nach den Zeitpunkten der Abgabe von Bewertungen sortiert. Für jede im Datensatz enthaltene Bewertung wurde nun auf dem simulierten Knoten, dessen Nummer der zu dieser Bewertung enthaltenen Nutzernummer entsprach, die Methode zum Hinzufügen einer Bewertung aufgerufen. Zugleich wurde auf dem Knoten der Zeitpunkt der Bewertung, wie er im Datensatz enthalten ist, mitgeteilt.

In den Simulationen wurden grundsätzlich die aggregierte Aktualisierung und aggregierte Speicherung von Bewertungen, Replikation von Objekttabellen und Multicast auf Anwendungsebene eingesetzt. Die in den Simulationen verwendeten Parameter sind in Tabelle 6.1 aufgeführt. Zur Evaluation einzelner Optimierungen wurden einzelne Parameter variiert; dies ist im jeweiligen Abschnitt beschrieben.

Die verwendeten Chord-IDs hatten eine Länge von 32 bit, die finger tables der Knoten 32 Einträge. Zudem kannte jeder Knoten seine 16 direkten Nachfolger.

Auf die redundante Datenspeicherung wurde in der Regel verzichtet; allerdings wurde in einzelnen Simulationsläufen bestätigt, dass der Kommunikationsaufwand in Parameter s linear und in Parameter r sublinear steigt. Soweit nicht anders erwähnt, wurde beim objektbasierten Ansatz für die aggregierte Aktualisierung der Parameter d aus Formel 5.14

[2] Bei den Simulationen zeigte sich, dass diese Skalierbarkeit zwar für den Aufbau eines Netzes besteht, der Nachrichtenaustausch zwischen den so erzeugten Knoten jedoch ineffizient gestaltet war; entsprechend mussten in Teilbereichen Anpassungen vorgenommen werden. Die Funktionalität selbst wurde dabei nicht geändert, jedoch das Format der verwendeten IDs angepasst sowie der Umfang der erzeugten Statistiken reduziert, wodurch der Speicheraufwand des Simulators erheblich verringert werden konnte.

sowie Parameter g aus Formel 5.16 auf $1,0$ gesetzt; als Parameter f aus Gleichung 5.14 wurden 90 Sekunden gewählt. Parameter c aus Gleichung 5.16 wurde auf den deutlich höheren Wert von einer Woche gesetzt. Der im Vergleich zum Parameter f deutlich höhere Wert liegt darin begründet, dass f beim Einfügen einer Bewertung herangezogen wird; bei einer zu großen Verzögerung besteht hier die Gefahr, dass der Knoten das Peer-to-Peer-System zwischenzeitlich verlässt. Bei dem durch den Parameter c mitbestimmten Abrufen aktualisierter Informationen besteht dieses Problem nicht.

Für den Empfehlungsalgorithmus des objektbasierten Ansatzes wurde α auf 0 gesetzt. Der Schwellenwert für die Zusammenfassung der Objekttabellen gemeinsam genutzter Objekte lag bei 100. Beim nutzerbasierten Ansatz fragte jeder Knoten vor dem Berechnen von Empfehlungen seine Nachbarknoten zwei Mal nach ihren taste buddies.

Die Ergebnisse einzelner Simulationsläufe können als Stichprobe aus der Gesamtheit aller möglichen Abläufe der Empfehlungserzeugung verstanden werden. Als Maß für den Stichprobenfehler wurde bei den durchgeführten Untersuchungen der Standardfehler σ_n herangezogen, der als Quotient aus Standardabweichung σ und der Wurzel aus der Anzahl gezogener Stichproben n definiert ist:

$$\sigma_n = \frac{\sigma}{\sqrt{n}} \tag{6.1}$$

Dieses Maß erlaubt die Einschätzung der Güte der jeweils ermittelten Mittelwerte; je kleiner der Standardfehler, desto zuverlässiger ist die Schätzung durch die gezogenen Stichproben.

Soweit nicht anders angegeben, wurden für jede Untersuchung 5 Simulationsläufe durchgeführt, wobei jeweils unterschiedliche Ausschnitte des Movielens-Datensatzes betrachtet (jeweils die Objekte mit den IDs 0–100, 100–200, ...) und das Chord-Netz neu initialisiert wurden. Trotz dieser geringen Zahl an Simulationsläufen konnten die auftretenden Stichprobenfehler klein gehalten werden: Ein durchschnittlicher Ausschnitt des Movielens-Datensatzes mit 100 Objekten beinhaltet bereits 25.309 Bewertungen, so dass eine geringe Varianz der in verschiedenen Simulationsläufen gemessenen Ergebnisse erzielt wurde.

6.2.3 Betrachtete Szenarien

Die durchgeführten Untersuchungen beruhen auf zwei Szenarien. In Szenario 1 wurden Bewertungen durch die Nutzer in der durch den Movielens-Datensatz vorgegebenen zeitlichen Reihenfolge abgegeben – der Datensatz für diesen Zweck nach Bewertungszeitpunkten sortiert –, bis der Datensatz vollständig eingelesen war. In Szenario 2 forderten nach jeweils 1000 eingelesenen Bewertungen alle Nutzer Objektempfehlungen an.

Um nach gewünschten Aufgaben des Empfehlungssystems differenzieren zu können, wurde weiterhin unterschieden, ob Nutzer relevante Objekte finden wollten (in diesem Fall wurden Precision und Recall als Metriken berechnet) oder auf die Vorhersage von Bewer-

Allgemeines	
Knotenzahl	10.000
Overlay/Chord	
Chord-IDs	32 bit
Finger table	32 Nachbarn
Bekannte direkte Nachfolger	16
Datenspeicherung	
r (vgl. Abschnitt 5.5)	1
s (vgl. Abschnitt 5.5)	1
Nutzerbasierter Ansatz	
taste buddies	40
z' (vgl. Abschnitt 5.6.5)	40
Durchläufe der taste-buddy-Suche	2
Wahrscheinlichkeit p (vgl. Abschnitt 5.6.5)	0
Kombinierter Ansatz	
γ (vgl. Abschnitt 5.6.6)	1
η (vgl. Abschnitt 5.6.6)	2
Weitere Parameter	Wie nutzerbasierter Ansatz
Objektbasierter Ansatz	
Empfehlungsalgorithmus:	
α (vgl. Abschnitt 5.8.2)	0
Replikation von Objekttabellen:	
l (vgl. Abschnitt 5.8.3.1, Gleichung 5.10)	0,2
Aggregierte Aktualisierung:	
c (vgl. Abschnitt 5.8.3.3, Gleichung 5.14)	604.800 s (1 Woche)
d (vgl. Abschnitt 5.8.3.3, Gleichung 5.14)	1,0
f (vgl. Abschnitt 5.8.3.3, Gleichung 5.16)	90 s
g (vgl. Abschnitt 5.8.3.3, Gleichung 5.16)	1,0
Aggregierte Speicherung:	
Zusammenfassung ab gemeinsamen Nutzungen	100

Tabelle 6.1: Parameter der Simulationen (vgl. Abschnitt 6.2.2)

tungen (mit MAE als Metrik; die Metriken werden im folgenden Abschnitt dargestellt) gezielt wurde.

6.2.4 Empfehlungsqualität

Die Messung der erzielten Empfehlungsqualität stellt sich als methodisch schwierig heraus; die Methodik und auftretende Schwierigkeiten werden in [HKTR04] ausführlich diskutiert. In einem ersten Schritt gilt es die Aufgaben des Empfehlungssystems zu identifizieren, die bewertet werden sollen. Für die vorliegende Arbeit werden die – auch in der Literatur am meisten betrachteten – Aufgaben diskutiert, die in Abschnitt 2.2.1 bereits eingeführt wurden: Bewertung von Objekten (in [HKTR04] allgemeiner als „Annotation“ bezeichnet) einerseits und die Empfehlung dem Nutzer noch unbekannter Objekte andererseits (in [HKTR04] als „find good items“ bezeichnet). Eine Reihe verwandter Aufgaben kann grundsätzlich ebenfalls mit den betrachteten Systemen erfüllt werden; diese sind jedoch nicht die Kernaufgaben, die bei den Entwurfsüberlegungen berücksichtigt wurden. Ohnehin ist ein Vergleich mit anderen Systemen bezüglich der Erfüllung anderer Aufgaben nur begrenzt möglich, da diese in der Literatur kaum berücksichtigt sind.

Im zweiten Schritt stellt sich die Frage, ob die Bewertung anhand von Experimenten mit realen Nutzern oder auf der Basis bereits vorhandener Datensätze („offline“) durchgeführt werden soll. Da neben der Bewertungsqualität auch die Skalierbarkeit gezeigt werden soll, würden sich Experimente mit realen Nutzern jedoch sehr aufwendig gestalten – insbesondere, wenn mehrere Systeme oder Algorithmen miteinander verglichen werden sollen (vgl. [HKTR04]). Zudem leidet unter der Auswahl der Testpersonen auch die Vergleichbarkeit mit Ansätzen aus der Literatur. Aus diesen Gründen wurde die Offline-Datenanalyse gewählt. Hierfür können frei verfügbare Testdatensätze verwendet werden, so dass Ergebnisse reproduzierbar und mit der Literatur – soweit der gleiche Datensatz verwendet wird – vergleichbar sind. Allerdings weist diese Form der Evaluation ebenfalls methodische Schwächen auf:

- Schlägt das System dem Nutzer ein Objekt vor, so kann dieser in einem realen Experiment entscheiden, ob er es für relevant hält oder nicht. Bei der Offline-Analyse aufgrund bestehender Datensätze kann die Tatsache, dass ein Nutzer ein Objekt nicht als relevant eingestuft oder nicht bewertet hat, auch schlicht dadurch begründet sein, dass er es nicht kennt. Dieser Effekt führt dazu, dass die Qualität von Objektvorschlägen systematisch unterschätzt wird (vgl. [HKTR04]).

- Die subjektive Einschätzung des Empfehlungssystems (beispielsweise aufgrund der Latenz oder anderer Kriterien) durch die Nutzer kann nicht bewertet werden [HKTR04].

- Für die Beurteilung der Skalierbarkeit des Systems ist auch der zeitliche Verlauf des Einfügens von Dokumenten in das System sowie des Abrufens von Empfehlungen von Bedeutung. Dieser zeitliche Verlauf ist jedoch nicht in allen gängigen Datensätzen enthalten.

Entscheidend ist neben der Auswahl der zu verwendenden Testdaten noch die Metrik, mit denen die Empfehlungsqualität auf diesen Daten gemessen wird. Die in der Evaluation verwendeten Metriken werden im folgenden Abschnitt vorgestellt.

6.2.4.1 Metriken

In [HKTR04] werden drei Klassen von Bewertungsmetriken unterschieden:

- Vorhersagegenauigkeitsmetriken (Predictive Accuracy Metrics), die die durch das System errechneten Bewertungen mit den von den Nutzern tatsächlich vergebenen vergleichen.
- Klassifikationsgenauigkeitsmetriken (Classification Accuracy Metrics), die die korrekte Einteilung von Objekten in Klassen (wie „relevant“, „nicht relevant“) bewerten.
- Ranggenauigkeitsmetriken (rank accuracy metrics), die die Genauigkeit messen, mit der ein Empfehlungssystem die Rangordnung von den Nutzern präferierter Objekte vorhersagt.

Im Rahmen dieser Arbeit werden Vertreter der ersten beiden Klassen gewählt:

- Der durchschnittliche absolute Fehler (Mean Absolute Error, MAE) als Vorhersagegenauigkeitsmetrik. Er ist definiert als die durchschnittliche Differenz zwischen der durch das Empfehlungssystem vorhergesagten und der durch den Nutzer anschließend tatsächlich vorgenommenen Bewertung:

$$MAE = \frac{\sum_{i=1}^{N} |E_{vorhergesagt} - E_{Nutzer}|}{N}, \tag{6.2}$$

  mit N Gesamtzahl an Bewertungen im System. Eine empirische Untersuchung [HKTR04] hat gezeigt, dass der MAE gut mit anderen Genauigkeitsmetriken korreliert, solange diese über alle Bewertungen im System und nicht pro Nutzer berechnet werden.[3] Er kann deshalb als Stellvertreter für diese Klasse von Metriken verwendet werden. Zu beachten gilt, dass der MAE keine Vergleichbarkeit zwischen verschiedenen Datensätzen herstellt. In [GRGP01] wird der normalisierte MAE vorgeschlagen, der sich als Quotient aus MAE und der Breite der verwendeten Bewertungsskala

[3] Dies gilt, obwohl der MAE selbst unabhängig davon ist, ob er pro Nutzer oder über alle Bewertungen des Systems errechnet wird.

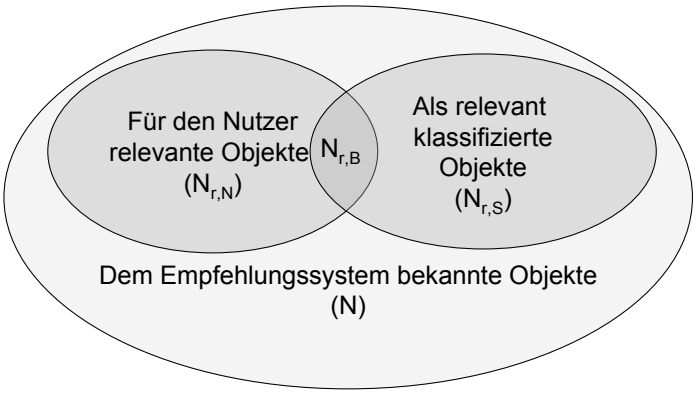

Abbildung 6.2: Precision und Recall

ergibt; ob sich dadurch die Vergleichbarkeit verschiedener Datensätze tatsächlich verbessert, ist jedoch bislang nicht empirisch untersucht oder theoretisch gezeigt worden.

- Precision und Recall [ClKe66] als Klassifikationsgenauigkeitsmetriken; sie messen die Genauigkeit der Einteilung in zwei Klassen („relevant“ und „nicht relevant“). Die Metriken stammen aus dem Bereich des Information Retrieval und werden für einzelne Nutzer berechnet, um anschließend den Durchschnitt über alle Nutzer zu bilden. Die Precision ist der Anteil tatsächlich als relevant empfundener Objekte an den als relevant klassifizierten Objekten (vgl. Abbildung 6.2.4.1).

$$Precision = \frac{N_{r,B}}{N_{r,S}} \tag{6.3}$$

Der Recall hingegen ist definiert als der Anteil der als relevant empfundenen und auch so klassifizierten Objekte an allen im System vorhandenen relevanten Objekten:

$$Recall = \frac{N_{r,B}}{N_{r,N}} \tag{6.4}$$

Sollen bei einem System, das Bewertungen von Objekten liefert, die über die Einteilung in zwei Klassen hinausgeht, Precision und Recall berechnet werden, so spielt die Schwelle, ab der Objekte als relevant eingestuft werden, eine entscheidende Rolle: Je höher diese angesetzt wird – je weniger Objekte dem Nutzer also tatsächlich empfohlen werden – desto geringer fällt typischerweise der Recall aus, wohingegen die Precision tendenziell steigt. Aus diesem Grund sind Precision und Recall auch nur in ihrer Kombination aussagekräftig. Die beiden Metriken wurden für die Evaluation im Rahmen der vorliegenden Arbeit aufgrund ihrer Anschaulichkeit ausgewählt.

Als Metrik für die Vorhersagegenauigkeit (für den Fall, dass Bewertungen vorhergesagt werden sollen), wird der oben bereits beschriebene Mean Absolute Error (MAE) herangezo-

gen. Betrachtet werden sowohl personalisierte Vorhersagen, also solche, die aufgrund der gemeinsamen Benutzungen mit Objekten aus dem eigenen Nutzerprofil entstehen (vgl. Abschnitt 5.8.2), als auch Objekte, für die das nicht möglich ist (hier wurde die Durchschnittsbewertung herangezogen). Konkret wurde der Mean Absolute Error dergestalt ermittelt, dass jeder Knoten für jedes Objekt, das im Movielens-Datensatz vorhanden und von diesem Knoten bewertet war, eine Empfehlung in Form einer vorhergesagten Bewertung anforderte – unabhängig davon, ob die Bewertung bereits in das System eingefügt war.

Bei der Empfehlungsberechnung hingegen wurden nur diejenigen Objekte berücksichtigt, die dem Nutzerprofil des jeweiligen Knotens zum entsprechenden Zeitpunkt bereits hinzugefügt worden waren – also diejenigen Bewertungen, deren Zeitstempel vor dem aktuell simulierten Zeitpunkt lag. Zudem wurde dabei für die Berechnung einer vorhergesagten Bewertung für ein bestimmtes Objekt eine eventuell bereits vorhandene Bewertung des jeweiligen Knotens für das entsprechende Objekt außen vor gelassen.

6.2.4.2 Objektbasierter Ansatz

In einem ersten Schritt wurde für den Ansatz des objektbasierten kollaborativen Filterns der Parameter α des verwendeten Ähnlichkeitsmaßes (Abschnitt 5.8.2) variiert. Die resultierende Empfehlungsqualität nach Ablauf des vollständigen simulierten Zeitraums und bei Einbeziehung aller Nutzer und Objekte, also unter Berücksichtigung aller im Movielens-Datensatz enthaltenen Bewertungen ist in Tabelle 6.2 dargestellt. Aufgrund der Verwendung des vollständigen Movielens-Datensatzes und vollständer Information aller Knoten wurde jeweils nur ein Simulationslauf durchgeführt; das Ergebnis ist deterministisch. Im besten Fall ergab sich ein MAE von 0,7848. Der Wert entspricht den in der Literatur erzielten Ergebnissen zentraler Empfehlungssysteme, die ebenfalls den Movielens-Million-Datensatz verwenden (so z.B. Xue et al. [XLYX+05], die Empfehlungen jedoch ausgehend von einem kleineren Ausschnitt dieses Datensatzes berechnen, sowie [O'He99]). Dieser Wert ist ausschließlich von der Parametrisierung des zugrunde gelegten Empfehlungsalgorithmus abhängig, da der objektbasierte Ansatz dem Algorithmus nach Einlesen aller Bewertungen vollständige Informationen über alle Bewertungen eines Objekts zur Verfügung stellen kann[4].

Werden statt des vollständigen Movielens-Datensatzes nur Ausschnitte mit jeweils 100 Objekten betrachtet und die Simulation mit den in Tabelle 6.1 dargestellten Parametern durchgeführt, so ergibt sich ein MAE von 0,801 bei einem Standardfehler von 0,027. Dabei konnten (bei einer Auswertung in Szenario 2) 0,561 Empfehlungen pro abgegebener Bewertung berechnet werden (Standardfehler 0,025).

[4] Diese Aussage gilt allerdings nur für einen idealisierten Ablauf – wenn alle Knoten erst Empfehlungen anfordern, nachdem alle im betrachteten Datensatz enthaltenen Bewertungen abgegeben sind, erhalten sie zu diesem Zeitpunkt vollständige Information. Das heißt jedoch nicht, dass im zeitlichen Verlauf ständig vollständige Informationen vorhanden sind.

α	0	0,1	0,2	0,3	0,4	0,5
MAE	0,7848	0,7849	0,7851	0,7852	0,7854	0,7857
α	0,6	0,7	0,8	0,90	1,0	
MAE	0,7859	0,7862	0,7866	0,7870	0,7874	

Tabelle 6.2: Empfehlungsqualität des objektbasierten Ansatzes

Die Werte, die für Precision und Recall erzielt wurden, hängen naturgemäß von der Anzahl empfohlener Objekte ab. Abbildung 6.3 zeigt, wie beide sich beim objektbasierten Ansatz in Abhängigkeit dieser Anzahl entwickeln. Dazu wurde wiederum der vollständige Movielens-Datensatz mit allen Benutzern herangezogen – hier mit vollständiger Information und dem Parameter $\alpha = 0$. Jeder Knoten hatte hier die Aufgabe, für den Benutzer relevante Objekte zu finden, die sodann nach ihren vorhergesagten Bewertungen in einer Rangliste sortiert wurden. Die n Objekte mit den besten Bewertungen wurden als relevant klassifiziert – der Wert für n ist dabei auf der Rechtsachse des Graphen aufgetragen. Lag für das Objekt tatsächlich eine Bewertung des jeweiligen Nutzers vor, die besser als 3 war, wurde es als tatsächlich relevant betrachtet. Mit dieser Parametrisierung wurde beispielsweise bei 20 empfohlenen Objekten ein Recall von 14,5% und eine Precision von 49,9% erreicht. Wie oben erwähnt, führt die Auswertung mit einem vorgefertigten Testdatensatz allerdings zu einem systematischen Unterschätzen beider Werte. Zu berücksichtigen ist ferner, dass die Definition eines Objekts als „relevant“ variiert werden kann. So wäre es auch möglich, diejenigen n Objekte als relevant zu bezeichnen, die ein Nutzer am besten bewertet hat; der Vergleich mit Werten aus der Literatur setzt also voraus, dass die Relevanz identisch definiert ist.

Der objektbasierte Ansatz wurde zusätzlich unter Verwendung des Bookcrossing-Datensatzes evaluiert. Herangezogen wurde dabei ein Ausschnitt dieses Datensatzes, nämlich Bewertungen über die ersten 30.000 Bücher, die darin enthalten sind – insgesamt wurden somit 142.638 Bewertungen berücksichtigt. Die Simulation beschränkt sich auf eine Evaluierung des Empfehlungsalgorithmus, dem in diesem Fall vollständige Informationen über alle Bewertungen vorlagen.

Es ergibt sich so rechnerisch ein MAE von $3,38$ (bei einer Bewertungsskala von 1–10); dieser Wert ist allerdings nur sehr bedingt aussagekräftig, da im Bookcrossing-Datensatz neben expliziten Bewertungen auch implizite Bewertungen enthalten sind, die lediglich Interesse an einem Objekt bedeuten. Bei Rückgabe von jeweils 20 empfohlenen Objekten wurde ein Recall von 51,9% und eine Precision von 42,9% erreicht, wobei ein Objekt dann als relevant klassifiziert wurde, wenn es eine Bewertung größer als 6 hatte. Zu beachten ist, dass in [ZMKL05] für einen objektbasierten Ansatz wesentlich kleinere Werte für Precision und Recall (3,64% bzw. 7,32%) erzielt werden. Dies liegt – neben der unterschiedlichen Wahl eines Ausschnitts des Bookcrossing-Datensatzes – an der dort verwendeten anderen Metho-

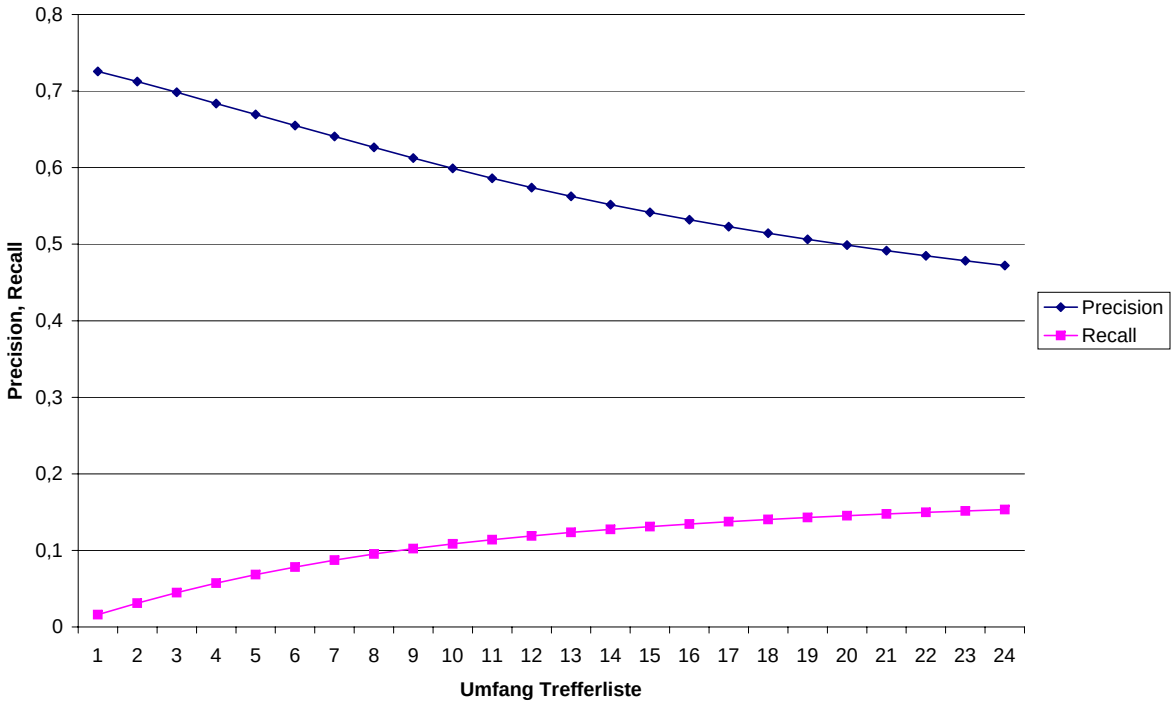

Abbildung 6.3: Precision und Recall beim objektbasierten Ansatz

dik zur Messung der Empfehlungsqualität. Insbesondere wurde in der vorliegenden Arbeit ein Objekt als relevant betrachtet, wenn der Nutzer es mindestens mit einer höheren Bewertungszahl als 6 (von 10) bewertet hatte; dem Nutzer wurden Objekte abhängig von ihrer vorhergesagten Bewertung vorgeschlagen. Dabei kann es auch vorkommen, dass dem Nutzer weniger als die genannten 20 Objekte vorgeschlagen werden. In [ZMKL05] haben die Ergebnislisten eine konstante Größe, auch wenn eigentlich nicht genügend relevante Objekte vorhanden sind; zudem wurden alle für einen Nutzer potentiell interessanten – also auch lediglich implizit bewertete – Objekte als relevant eingestuft.

6.2.4.3 Nutzerbasierter Ansatz

Beim nutzerbasierten Ansatz steht zu erwarten, dass die Empfehlungsqualität und die Anzahl von Empfehlungen, die berechnet werden können, von t abhängig ist. Dies wurde auch in der simulativen Evaluierung bestätigt. Abbildung 6.5 zeigt die Anzahl an Empfehlungen, die erzeugt werden konnten und Abbildung 6.6 den erzielten MAE in dem Szenario aus Tabelle 6.1, bei dem wiederum Ausschnitte des Movielens-Datensatzes mit 100 Objekten betrachtet wurde (jeweils mit Standardfehler). Dabei wurden jeweils zwei Durchläufe der Suche nach ähnlichen Knoten durchgeführt. Die Anzahl der erzeugten Empfehlungen wurde auf die Anzahl der jeweils eingefügten Bewertungen normiert.

Wie sich herausstellt, bringt in diesem Szenario eine Steigerung der betrachteten Nach-

barschaft bis hin zu ca. 70 Knoten einen deutlichen Vorteil sowohl bezüglich der erreichten Empfehlungsqualität als auch der Anzahl berechneter Empfehlungen. Allein die Verdopplung der Zahl der taste buddies von 10 auf 20 bringt dabei eine Erhöhung der Empfehlungszahl um über 50%. Diese Beobachtung ist in guter Übereinstimmung mit theoretischen Erwartungen. Abbildung 6.4 zeigt die kumulative Verteilungsfunktion einer hypergeometrischen Verteilung; diese Verteilung wird in der Statistik benutzt, um die Wahrscheinlichkeit auszudrücken, dass in einer zufällig (ohne Zurücklegen) gezogenen Stichprobe eine bestimmte Anzahl von Elementen mit einer Eigenschaft vorkommt [Schl98]. Die hypergeometrische Verteilung lässt sich durch drei Parameter beschreiben: N ist die Anzahl der Elemente der Grundgesamtheit, aus der die Stichprobe gezogen wird; entsprechend der Nutzerzahl im Movielens-Datensatz wurde N auf 6.040 gesetzt. M ist die Anzahl der Elemente mit der gewünschten Eigenschaft – hier das Vorhandensein eines bestimmten Objekts im Nutzerprofil eines Nutzers. Im Mittel liegt jedes Objekt, das im Movielens-Datensatz enthalten ist, in den Nutzerprofilen von 253 Nutzern vor; dieser Wert wurde also für M gewählt. n ist die Anzahl der Elemente einer Stichprobe. Die hypergeometrische Verteilung liefert dann für ein gegebenes k die Wahrscheinlichkeit, dass in der Stichprobe k Elemente mit der gewählten Eigenschaft vorhanden sind.

$$h(k|N;M;n) := P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \tag{6.5}$$

Die Auswahl der taste buddies lässt sich nun als Ziehen einer Stichprobe aus der Grundgesamtheit der Nutzer verstehen; jeder taste buddy hat die Eigenschaft, ein bestimmtes Objekt bewertet zu haben oder nicht. In Abbildung 6.4 ist nun

$$1 - h(0|6040;253;n)$$

in Abhängigkeit von n aufgetragen – also die Wahrscheinlichkeit, dass von n zufällig gewählten taste buddies mindestens einer ein Objekt bewertet hat, das insgesamt durchschnittlich oft bewertet wurde.

Die Emfehlungsqualität bleibt – wie in Abbildung 6.6 gezeigt – deutlich unter dem Wert des objektbasierten Ansatzes: Bestenfalls wird ein MAE von $0,945$ erreicht – mit dem objektbasierten Ansatz ließ sich im gleichen Szenario bei jeder verwendeten Parametrisierung ein Wert deutlich unter $0,9$ erzielen (vgl. dazu Abschnitt 6.2.4.2). Bestätigt wird dieser Vergleich bei Betrachtung von Precision und Recall (Abbildung 6.7). So konnte bei 20 empfohlenen Objekten lediglich eine Precision von 23,6% und ein Recall von 27,0% erzielt werden.

Die erzielten Werte sind ähnlich zu denen, die mit nicht-personalisierten Vorhersagen durch Verwendung der jeweils durchschnittlichen Bewertung eines Objekts erreichbar sind.

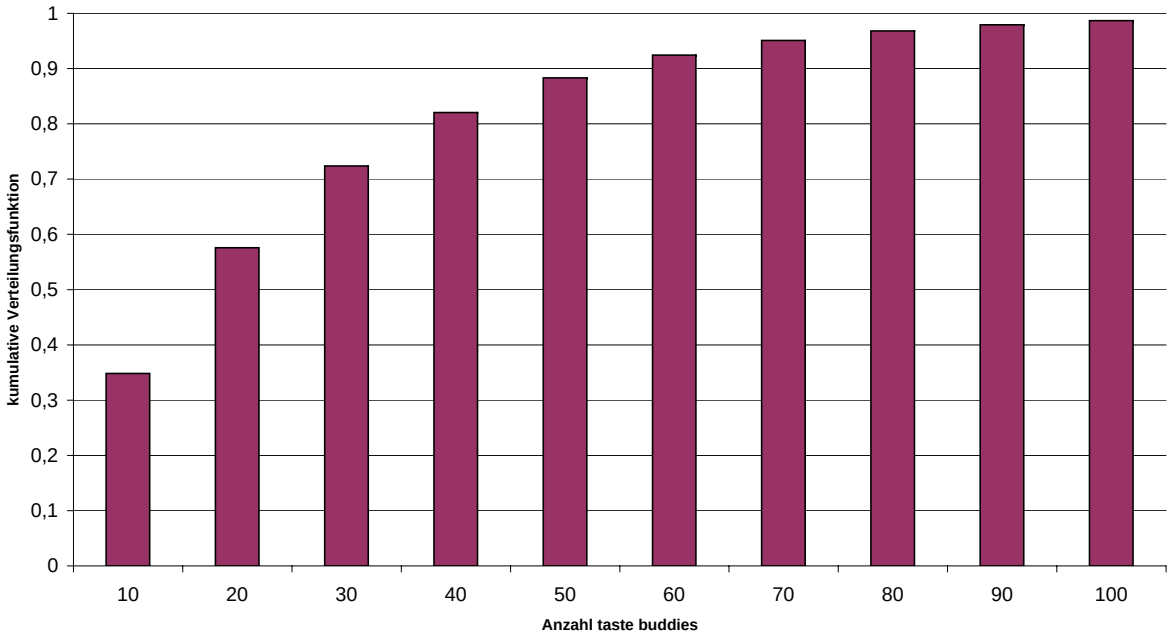

Abbildung 6.4: Hypergeometrische Verteilung

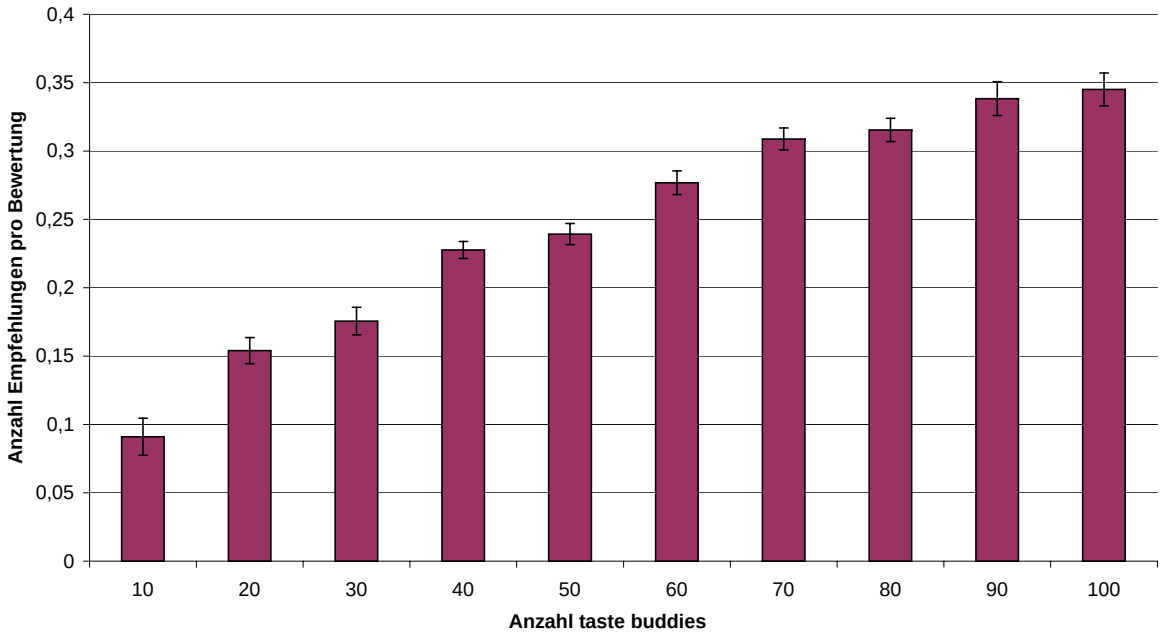

Abbildung 6.5: Anzahl der Empfehlungen im nutzerbasierten Ansatz

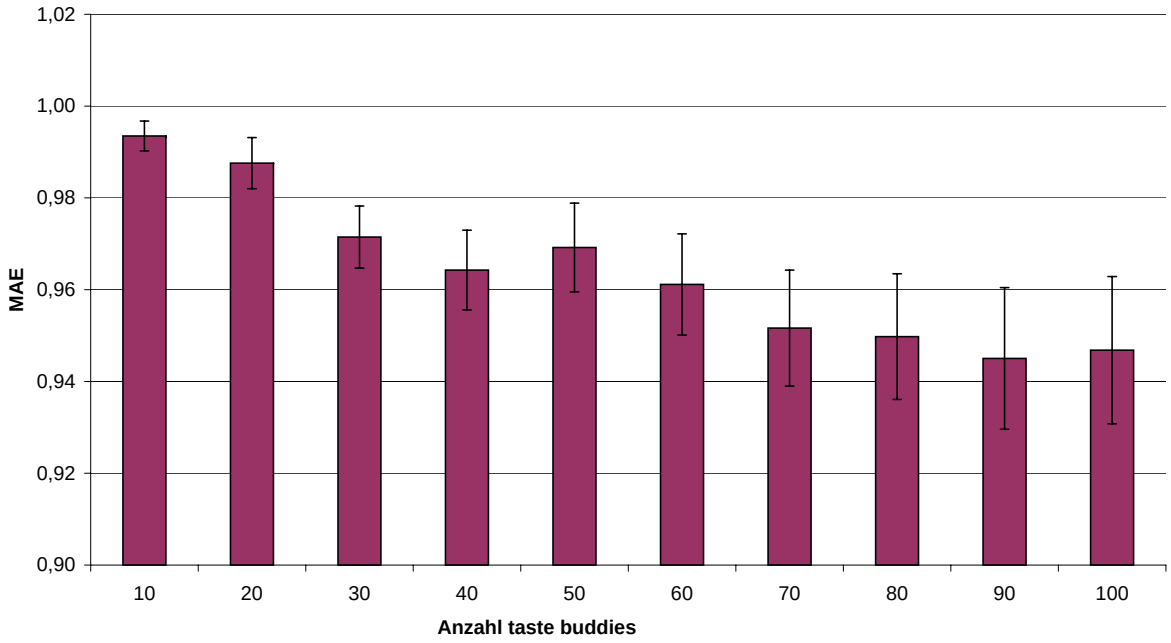

Abbildung 6.6: MAE im nutzerbasierten Ansatz

Durch Erhöhung der taste-buddy-Zahl kann nur eine geringfügige Verbesserung erreicht werden. Da beim nutzerbasierten Ansatz die Berechnung der durchschnittlichen Bewertung eines Objekts durch alle anderen Nutzer nicht möglich ist – es ist jeweils nur eine kleine Anzahl von Nutzerprofilen verfügbar – heißt das nicht, dass auf den verwendeten personalisierten Empfehlungsalgorithmus verzichtet werden kann.

Auch die Zahl pro Bewertung erzeugter Empfehlungen bleibt selbst bei 100 taste buddies mit 0,345 unter dem Wert von 0,561 des objektbasierten Ansatzes (vgl. Abbildung 6.5).

6.2.4.4 Kombinierter Ansatz

Der kombinierte Ansatz verbessert den nutzerbasierten Ansatz wie folgt: Wird auf einem Knoten eine Bewertung für ein Objekt hinzugefügt, das in dessen Nachbarschaft weniger als γ mal vorhanden ist, so wird ein Verweis auf den benutzenden Knoten unter dem Schlüssel des Objekts gespeichert. Dabei werden jedoch maximal η Verweise pro Objekt vorgehalten (vgl. Abschnitt 5.6.6).

Für die Evaluierung wurde γ auf den Wert 1 gesetzt. Ein Knoten, der eine Bewertung abgibt, speichert also nur dann einen Verweis auf das zugehörige Bewertungsdokument unter dem Schlüssel des Objekts ab, wenn dieses in seiner Nachbarschaft noch überhaupt nicht bekannt ist. Wird γ höher gewählt, erhöht sich der Kommunikationsaufwand, da dann auch Verweise auf in der Nachbarschaft bereits bekannte Objekte gespeichert werden. Deshalb soll untersucht werden, ob bereits bei diesem Wert eine hohe Abdeckung an Objekten

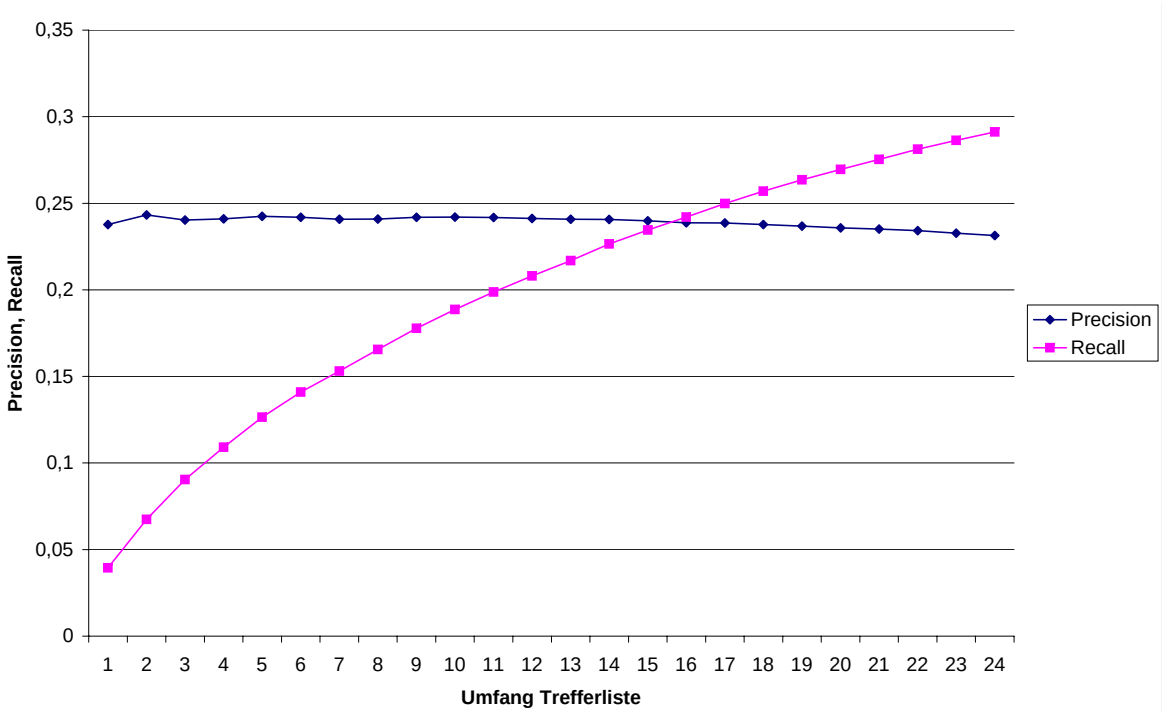

Abbildung 6.7: Precision und Recall im nutzerbasierten Ansatz

erreicht werden kann, deren Empfehlung vorhergesagt werden kann. Für η wurde der Wert 2 gewählt – es werden also pro Objekt höchstens zwei Verweise gespeichert.

Mit diesem Ansatz wurde eine Empfehlungsqualität erzielt, die zwischen derjenigen des nutzerbasierten und der des objektbasierten Ansatzes liegt. Abbildung 6.8 zeigt den MAE des kombinierten Verfahrens in Abhängigkeit von der Anzahl der taste buddies. Der Einfluss der Anzahl an taste buddies auf die Empfehlungsqualität ist geringer als beim rein nutzerbasierten Verfahren; ein Trend, nach dem mehr taste buddies zu einer besseren Empfehlungsqualität führen, kann nicht mehr ausgemacht werden. Bei wenigen taste buddies ist zwar die Wahrscheinlichkeit gering, dass unter den eigenen taste buddies eine hinreichende Anzahl an Bewertungen für ein bestimmtes Objekt vorhanden ist, um eine zutreffende Empfehlung zu erhalten. Ist ein Objekt von den eigenen taste buddies aber überhaupt nicht bewertet worden, kann der kombinierte Ansatz meist dennoch mehrere Bewertungen finden.

Für die Anzahl erzeugter Empfehlungen gilt das Gleiche: Auch hier lässt sich kein eindeutiger Trend ausmachen, wonach mehr taste buddies zu einer höheren Anzahl erzeugter Empfehlungen führen würden. Dies ist in Abbildung 6.9 dargestellt. Schon bei nur 10 taste buddies werden 0,47 Empfehlungen pro abgegebener Bewertung erzeugt – ein Wert, der knapp unter dem objektbasierten Ansatz liegt, aber deutlich über dem nutzerbasier-

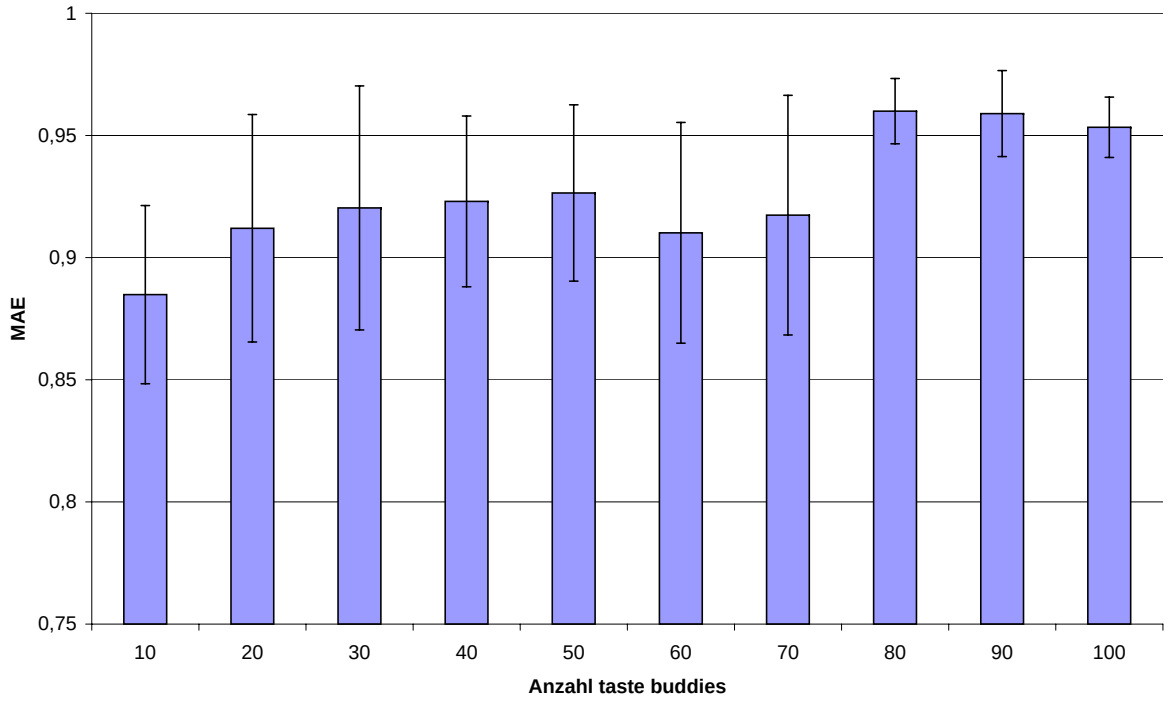

Abbildung 6.8: MAE des kombinierten Ansatzes (mit Standardfehler)

ten Ansatz. Der Wert wird allerdings auch bei größeren taste-buddy-Zahlen zunächst nicht überschritten – dies ist erst bei 90 taste buddies wieder der Fall.

Im Ergebnis kann also bereits mit sehr wenigen taste buddies eine hohe Wahrscheinlichkeit erreicht werden, dass für ein gegebenes Objekt eine Empfehlung berechnet werden kann. Dies bedeutet einen klaren Vorteil gegenüber dem nutzerbasierten Ansatz.

6.2.5 Skalierbarkeit

Im Folgenden wird die Skalierbarkeit der drei untersuchten Ansätze betrachtet.

6.2.5.1 Objektbasierter Ansatz

Bezüglich der Skalierbarkeit gestaltet sich beim objektbasierten Ansatz das Einfügen neuer Bewertungen in das System am kritischsten. Hier ist im theoretisch schlechtesten Fall – also dann, wenn die Zusammenfassung abgegebener Bewertungen scheitert – mit einem Kommunikationsaufwand zu rechnen, der quadratisch mit der Anzahl von Einträgen in den Profilen der Nutzer wächst: In diesem Fall muss bei jedem Einfügen einer Bewertung eine Nachricht an alle Knoten gesendet werden, die für vorher genutzte Objekte zuständig sind. Für eine einzelne Bewertung führt dies zu einem Kommunikationsaufwand von $O(u \log n)$ (u Anzahl Einträge im Nutzerprofil des Knotens, n Anzahl der Teilnehmer im System; der

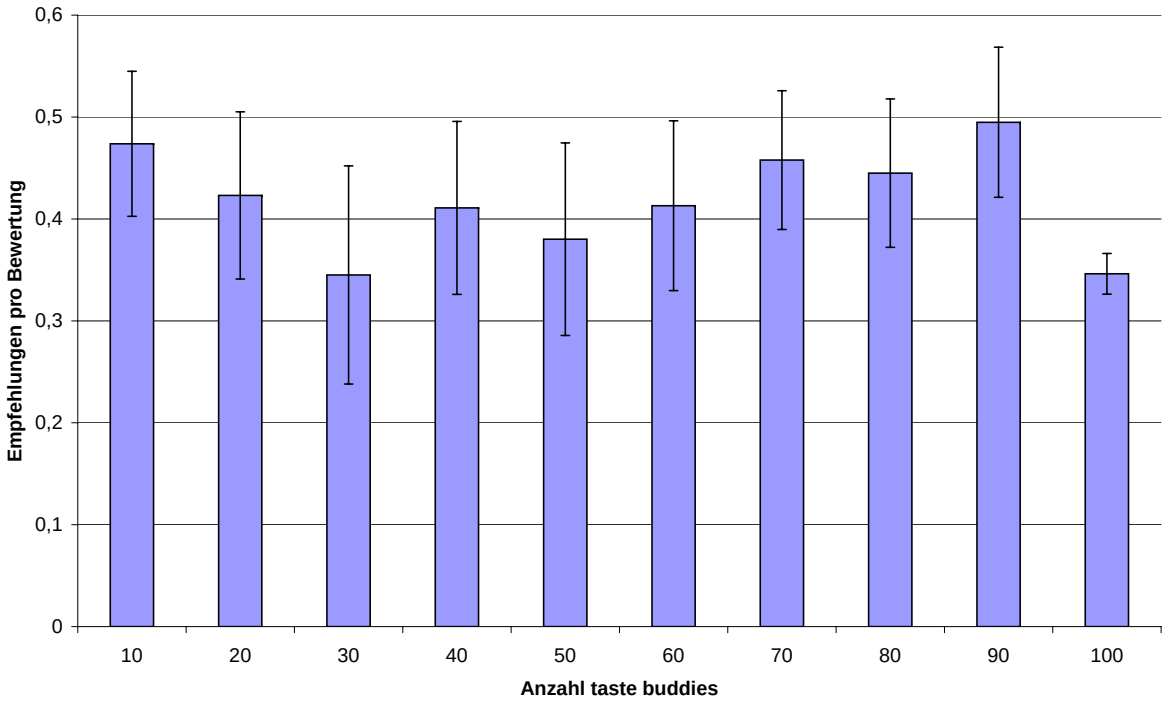

Abbildung 6.9: Anzahl erzeugter Empfehlungen, kombinierter Ansatz

Faktor $\log n$ ergibt sich aus dem Kommunikationsaufwand für das Chord-Routing). Quadratisch wird der Aufwand dadurch, dass der Prozess für jeden neuen Eintrag im Nutzerprofil wiederholt werden muss. Er liegt dann in $O(u^2 logn)$ (vgl. Abschnitt 5.8).

Die Optimierungen, die in Abschnitt 5.8.3 beschrieben werden, führen jedoch – zumindest in der Größenordnung des MovieLens-Datensatzes – zu einem praktisch noch handhabbaren Aufwand. Einen wesentlichen Beitrag dazu leistet die Tatsache, dass Nutzer – wie auch im Movielens-Datensatz ersichtlich – dazu neigen, mehrere Bewertungen in enger zeitlicher Folge (im Bereich weniger Minuten) abzugeben, was eine Zusammenfassung der entsprechenden Aktualisierungsnachrichten ermöglicht.

Die Abhängigkeit zwischen der Anzahl von Einträgen in den Nutzerprofilen und dem Aufwand für das Einfügen neuer Bewertungen wurde abgebildet, indem verschieden große Ausschnitte des Movielens-Datensatzes betrachtet wurden. Bei konstanter Nutzerzahl wurde die Anzahl bewerteter Objekte – und somit die durchschnittliche Größe eines Nutzerprofils – variiert. So wurden zunächst die jeweils ersten 100–900 Objekte ausgewählt und der betrachtete Ausschnitt dann jeweils um 900 Objekte verschoben; so wurde eine Überschneidung der jeweils verwendeten Teildatensätze vermieden. Allerdings standen aufgrund der Größe des Movielens-Datensatzes nur vier solcher Ausschnitte zur Verfügung. Es wurde Szenario 2 betrachtet; alle Knoten forderten also nach jeweils 1000 eingefügten Bewertungen Empfehlungen auf Basis ihres Nutzerprofils an. Verglichen wurde nun die Anzahl versand-

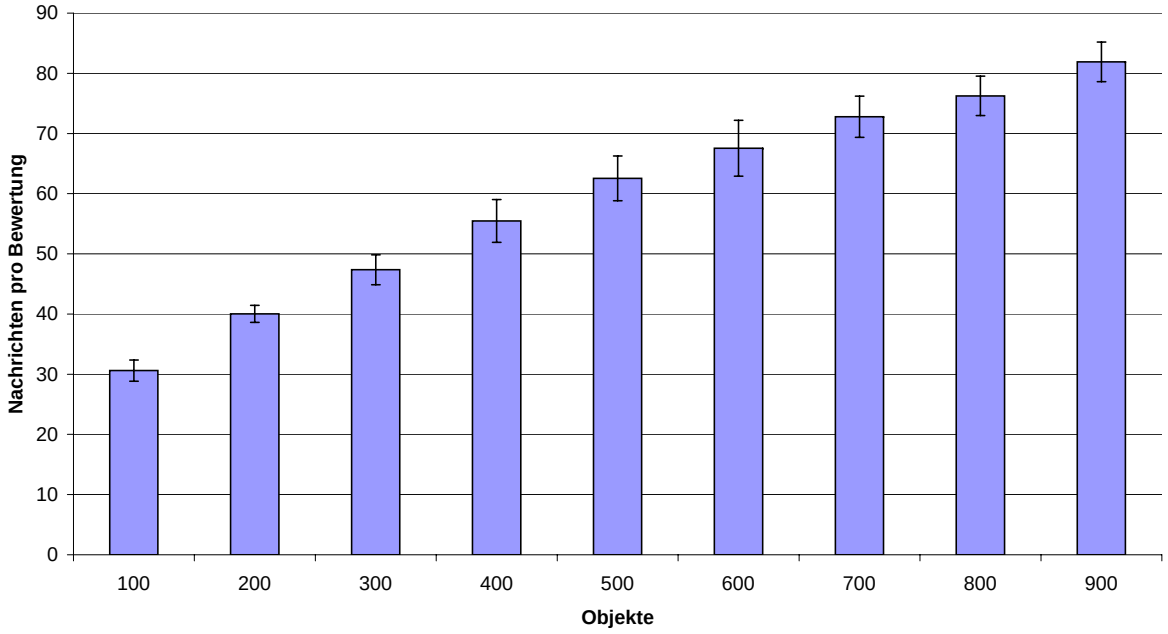

Abbildung 6.10: Versandte Nachrichten pro eingefügter Bewertung

ter Nachrichten auf Transportschicht, die jeweils im Durchschnitt über den gesamten zeitlichen Verlauf für die Erzeugung einer Empfehlung sowie für das Einfügen einer Bewertung benötigt wurden. Die Nachrichtenanzahl für die Erzeugung einer Empfehlung blieb dabei konstant bei 11,2 (bei einer Standardabweichung von 1,2). Mehrere dieser Nachrichten parallel zu versenden, ist durch das Anwendungs-Multicast-Verfahren nicht möglich, so dass sich aus diesen 11,2 Nachrichten auch die Latenz für das Erzeugen einer Empfehlung ergibt. Zu welcher Latenz dies tatsächlich führt, hängt von den Charakteristika des Underlay ab, die im PlanetSim-Simulator nicht modelliert werden können – geht man exemplarisch von 200 ms für den Versand einer Nachricht auf der Transportschicht aus, liegt diese Latenz beispielsweise bei 2,2 Sekunden.

Erwartungsgemäß galt der konstante Aufwand jedoch nicht für das Einfügen einer Bewertung; bei 100 Objekten lag der Aufwand hierfür bei 31,7 Nachrichten, um sich bei 400 Objekten auf 63,8 Nachrichten zu verdoppeln und bei 900 Objekten auf 81,9 zu wachsen. Die Anzahl pro Bewertung verschickter Nachrichten steigt also zwar mit der Anzahl im System vorhandener Objekte an, aber weniger als linear (vgl. Abbildung 6.10), da jeweils mehrere eingefügte Bewertungen zusammengefasst werden können. Der Kommunikationsaufwand wächst somit nicht mehr quadratisch mit der Anzahl im System vorhandener Objekte – mithin auch nicht mit der Anzahl in den Nutzerprofil enthaltener Objekte, die aufgrund der konstant gehaltenen Nutzerzahl der Gesamtzahl der Objekte proportional ist. Aus den Simulationsdaten ergibt sich ein Aufwand für das Einfügen von Bewertungen, der in $O(u^{1,45})$ liegt (Standardfehler $0,05$).

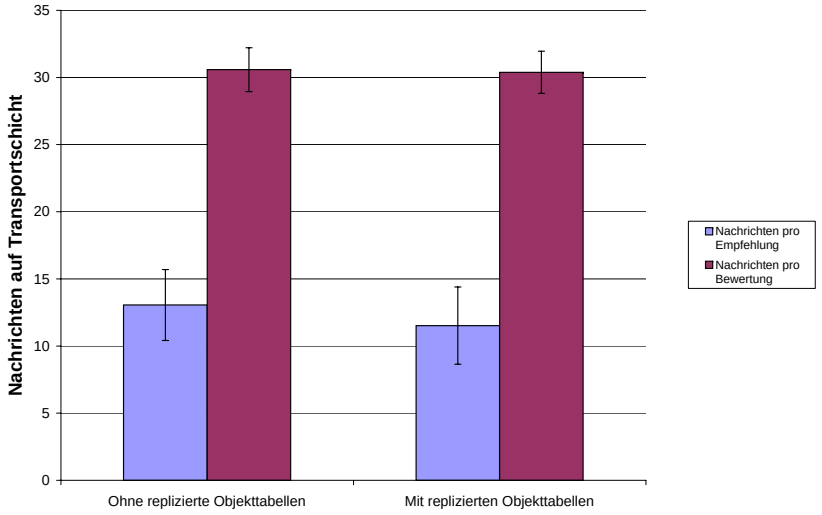

Abbildung 6.11: Replikation von Objekttabellen (mit Standardfehler)

Weiterhin wurden Simulationen durchgeführt, um die Wirksamkeit der einzelnen durchgeführten Optimierungen zu evaluieren.

Replikation von Objekttabellen Die Replikation von Objekttabellen dient einerseits der Lastverteilung, andererseits der Verkürzung der Pfadlängen, die im Chord-Ring beim Erstellen von Empfehlungen überwunden werden müssen. Bei der Simulation wurden die Parameter aus Tabelle 6.1 verwendet und wiederum lediglich 100 Objekte aus dem Movielens-Datensatz ausgewählt. Der Aufwand pro eingefügter Bewertung ergibt sich aus einem Simulationslauf, bei dem keine Empfehlungen erzeugt wurden; die Anzahl versandter Nachrichten wurde durch die Anzahl eingefügter Bewertungen geteilt. Sodann wurde ein identisch parametrisierter Simulationslauf durchgeführt, bei dem nach jeweils 1000 eingefügten Bewertungen Empfehlungen generiert wurden (Szenario 2). Die Differenz der in beiden Simulationsläufen Anzahl versandter Nachrichten, geteilt durch die Anzahl erzeugter Empfehlungen, ergibt den Aufwand pro Empfehlung. Beide Simulationsläufe wurden mit verschiedenen Teilmengen des Movielens-Datensatzes (Objekt-IDs 1 – 100, 101 – 200) wiederholt.

Wie in Abbildung 6.11 ersichtlich, kann der Aufwand für das Erzeugen einer Empfehlung leicht gesenkt werden; das Einfügen einer Bewertung erfordert hingegen geringfügig erhöhten Kommunikationsaufwand. Dies liegt darin begründet, dass beim Einfügen von Bewertungen der Replikationsprozess angestoßen wird, der selbst Kommunikationsaufwand verursacht.

Multicast auf Anwendungsebene Wie oben diskutiert, bedeutet diese Funktionalität, dass Nachrichten, die an mehrere Empfänger innerhalb des Chord-Rings geroutet werden, nicht

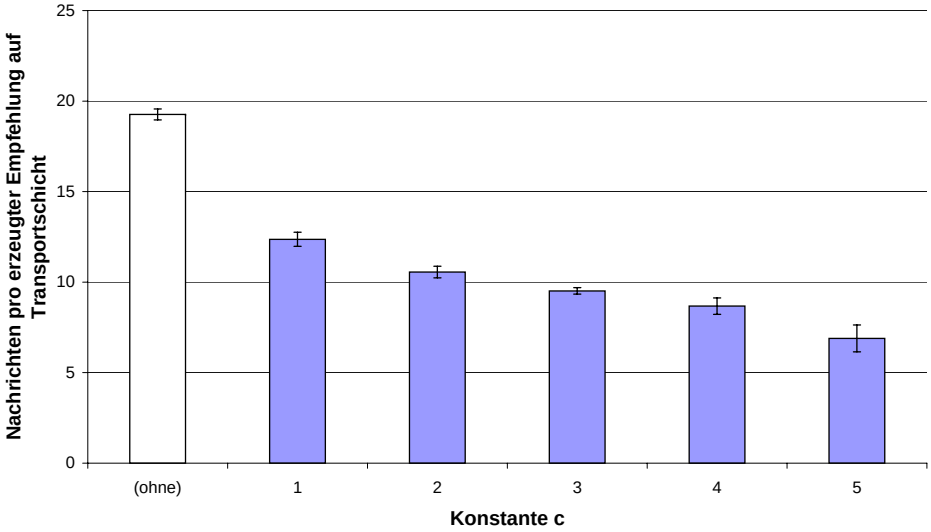

Abbildung 6.12: Nachrichten pro erzeugter Empfehlung (mit Standardfehler) mit und ohne aggregierte Aktualisierung

als einzelne Nachrichten verschickt, sondern innerhalb des Rings von einem Empfänger zum jeweils nächsten weitergeleitet werden. Im betrachteten Szenario 2 mit jeweils 100 betrachteten Objekten konnte die Anzahl an Nachrichten auf Transportschicht – bei gleichbleibender Nachrichtenzahl im Overlay – auf diesem Weg um 16,1% reduziert werden (Standardfehler 1,2%).

Aggregierte Aktualisierung Die aggregierte Aktualisierung von Bewertungen, die in die Empfehlungsberechnung einfließen, verringert den Aufwand zur Erzeugung einer Empfehlung erheblich. Die Untersuchung wurde anhand von Szenario 2 durchgeführt, es wurden also kontinuierlich Empfehlungen erzeugt. Wiederum liegen 100 Objekte aus dem Movielens-Datensatz zugrunde. Abbildung 6.12 zeigt die Anzahl an Nachrichten, die auf Transportschicht pro erzeugter Empfehlung versandt werden, in Abhängigkeit von der Konstanten c aus Formel 5.14. Je höher der Wert c ist, desto mehr Nachrichten können bei der Aktualisierung zusammengefasst werden. Zu erwarten ist dabei auch eine Verringerung der Empfehlungsqualität, weil nicht zu jedem Zeitpunkt aktuelle Informationen über gemeinsame Objektbenutzungen zur Verfügung stehen.

Im Mittel über alle während des Simulationslaufs erzeugten Empfehlungen zeigt sich im Vergleich der beiden in Abbildung 6.12 dargestellten Extremfälle (ohne aggregierte Aktualisierung und $c = 5,0$) aber kein nennenswerter Unterschied der erzielten Empfehlungsqualität; der MAE verschlechtert sich lediglich von 0,782 auf 0,794. Deutlicher wird der Unterschied allerdings während der Initialisierungsphase des Systems: Nach Einfügen von 5000

Bewertungen wird für $c = 5$ ein MAE von 0,824 erzielt; für $c = 0,01$ liegt dieser Wert bei 0,791.

Aggregierte Speicherung von Bewertungen Die aggregierte Speicherung von Bewertungen dient dem Ziel, den Aufwand für die Aktualisierung der lokalen Objekttabellen des Knotens zu reduzieren, der eine Empfehlung berechnen will. Dazu werden zwei Objekttabellen, die unter verschiedenen Schlüsseln gespeichert werden, zusammengefasst, sobald ein Schwellenwert an gemeinsamen Benutzungen erreicht wird. Hierbei entsteht jedoch auch zusätzlicher Kommunikationsaufwand:

- Sobald der Schwellenwert überschritten wird, ab dem Objekttabellen zusammengefasst werden, muss eine Tabelle zum neu zuständigen Knoten übermittelt (d.h. unter dem Schlüssel des anderen Objekts gespeichert) werden.
- Beim erstmaligen Abruf einer Objekttabelle, die nicht unter ihrem ursprünglichen Schlüssel gespeichert ist, muss dem nachfragenden Knoten der neue Schlüssel mitgeteilt und der Abruf erneut gestartet werden.

Bei wiederholten Abrufen oder Aktualisieren der Objekttabelle sollte der durchschnittliche Kommunikationsaufwand sinken: Jeder Knoten, in dessen Nutzerprofil beide Objekte vorhanden sind, spart bei einer Aktualisierung der lokalen Kopien der Objekttabellen den Abruf einer Tabelle unter einem weiteren Schlüssel. Der Effekt konnte jedoch simulativ nicht nachgewiesen werden. Im betrachteten Szenario 1 mit 100 Objekten aus dem Movielens-Datensatz sinkt zwar die Nachrichtenzahl für das Einfügen einer Bewertung bei einer aggregierten Speicherung ab 25 gemeinsamen Benutzungen von 29,1 auf 28,7. Die Differenz liegt aber noch innerhalb des Standardfehlers von 0,70 bzw. 0,71.

Begründen lässt sich dies durch die geringe Anzahl an Objekten, deren Objekttabellen zusammengefasst werden. Dies ist beabsichtigt; die aggregierte Speicherung der Objekttabellen soll lediglich besonders häufig kombinierte Objekte zusammenfassen. Ein messbarer Effekt ist auf diesem Weg allerdings erst nach einer langen Laufzeit zu erwarten, da der Mehraufwand einmalig entsteht und bei jeder eingefügten Bewertung oder berechneten Empfehlung nur eine geringe Einsparung erzielt wird. Um dies festzustellen, müssten weitere Datensätze herangezogen werden.

6.2.5.2 Nutzerbasierter Ansatz

In der Basisversion des nutzerbasierten Ansatzes werden Bewertungen eines Nutzers (also die Zeilen der Nutzer-Objekt-Matrix aus Abbildung 2.4) unter dem Schlüssel des jeweiligen Nutzers gespeichert. Für jedes Einfügen einer Bewertung sind also $O(\log N)$ Nachrichten nötig (N Anzahl der Knoten im Peer-to-Peer-System). Beim Eintritt in das System sucht

zudem jeder Knoten sich eine gewisse Anzahl t an sogenannten taste buddies. Zunächst werden dabei Profile herangezogen, die in der Nachbarschaft des jeweiligen Knotens gespeichert sind; der Suchaufwand ist also unabhängig von der Gesamtzahl der Knoten im Peer-to-Peer-System. Auch in weiteren Schritten vergleicht ein Knoten das eigene Profil lediglich mit den taste buddies der bereits bekannten taste buddies. Der Kommunikationsaufwand hierfür hängt also nur von der Größe der Liste der taste buddies ab und liegt in $O(t)$. Zum Erzeugen einer Empfehlung muss ein Knoten wiederum die Profile seiner taste buddies abrufen; auch hier liegt der Aufwand in $O(t)$. Von der Anzahl an Objekten in den Nutzerprofilen ist der Aufwand nicht abhängig. Pro taste buddy müssen für den eigentlichen Prozess des Berechnens von Empfehlungen insgesamt durchschnittlich 5,1 Nachrichten versandt werden; dies setzt sich zusammen aus 4,1 Nachrichten für das Chord-Routing auf dem Weg zum taste buddy und einer Antwortnachricht und gilt unabhängig von der Anzahl zu einem Zeitpunkt zu berechnender Empfehlungen. Der Prozess kann für mehrere taste buddies parallelisiert werden, so dass sich die zu erwartende Latenz für die Empfehlungsberechnung lediglich aus diesen 5,1 Nachrichten ergibt – bei beispielsweise 200 ms für eine Transportschicht-Nachricht als ca. 1 Sekunde.

Das Auffinden ähnlicher Knoten führt allerdings zu einem beträchtlichen Kommunikationsaufwand. Jeder Knoten muss seine taste buddies suchen und die Suchanfragen anderer Knoten beantworten. Bei 50 taste buddies entsteht so pro Knoten (in der Basisversion der verwendeten Parametrisierung gemäß Tabelle 6.1) ein durchschnittlicher Aufwand von 2844 versandten Nachrichten. Allerdings muss dieser Prozess für jeden Knoten nur einmal durchgeführt werden. Eine mehrmalige Suche nach neuen taste buddies ist zwar grundsätzlich vorgesehen. Da jeweils nur Knoten hinzugefügt werden, die ähnlicher sind als die vorher bereits vorhandenen taste buddies, sollte dies einen positiven Effekt auf die erzielbare Empfehlungsqualität haben. Es zeigt sich in den Simulationen jedoch, dass die tatsächlich erzielte Empfehlungsqualität auf diesem Weg nur geringfügig verbessert werden kann. Im Durchschnitt sank der MAE durch eine zusätzliche Suche nach neuen taste buddies lediglich um 0,0073 (Standardfehler 0,0031).

Ähnliches gilt für das zufällige Weiterleiten von Suchnachrichten bei der taste-buddy-Suche. In den Experimenten wurde der Parameter p aus Abschnitt 5.6.5 auf $0,1$ gesetzt und ansonsten die Parametrisierung aus Tabelle 6.1 beibehalten. Zwar stieg der Kommunikationsaufwand dabei über die gesamte Laufzeit betrachtet nur um 3,7% – ein Effekt auf die Empfehlungsqualität konnte aber nicht nachgewiesen werden.

6.2.5.3 Kombinierter Ansatz

Abbildung 6.13 zeigt exemplarisch für ein spezifisches Szenario (Szenario 1, Bewertungen der ersten 100 Objekte des Movielens-Datensatzes, Parametrisierung nach Tabelle 6.1) die absolute Anzahl von Nachrichten auf Transportschicht, die im zeitlichen Verlauf eingefügt

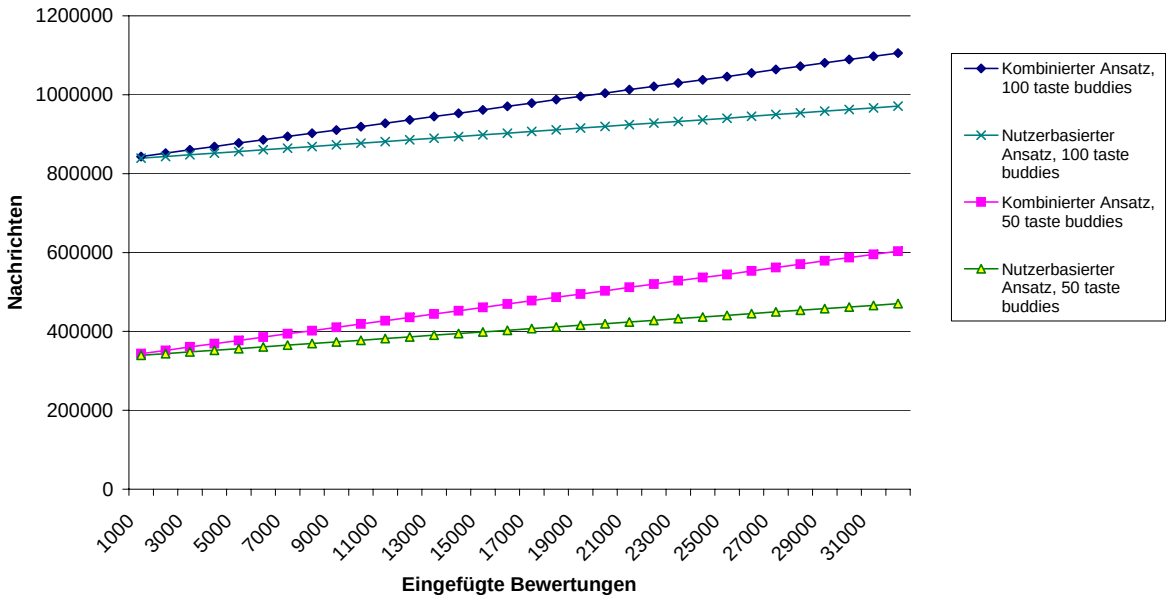

Abbildung 6.13: Einfügen von Bewertungen, nutzerbasierter und kombinierter Ansatz

werden. Im Mittel über die wiederum durchgeführten fünf Simulationsläufe mit verschiedenen Ausschnitten des Movielens-Datensatzes entstand ein Mehraufwand von 4,17 Nachrichten pro eingefügter Bewertung (10 taste buddies, Standardfehler 0,0002) bzw. 3,63 Nachrichten (80 taste buddies, Standardfehler 0,57). Mit steigender Anzahl an taste buddies sinkt dieser Mehraufwand, da die Wahrscheinlichkeit steigt, dass ein Knoten, der eine Bewertung einfügt, weitere Bewertungen des Objekts bereits in seiner Nachbarschaft vorfindet.

Jeder Knoten versuchte, für jedes ihm bekannte Objekt eine Empfehlung zu berechnen – auch, wenn dieses in der Nachbarschaft des Knotens nicht bekannt war. Hierbei wurde das gleiche Szenario wie bereits in der Basisversion des nutzerbasierten Ansatzes betrachtet. Es konnten bereits bei 10 taste buddies 0,42 Empfehlungen[5] pro abgegebener Bewertung erreicht werden – ein Wert, der bei steigender Anzahl an taste buddies zunächst sinkt und erst bei 70 taste buddies mit 0,39 wieder erreicht wird. Der rein nutzerbasierte Ansatz erreichte hingegen mit 10 taste buddies lediglich $0,091$ Empfehlungen pro eingefügter Bewertung (bei einem Standardfehler von 0,0033; vgl. Abbildung 6.5).

Die Berechnung einer Empfehlung läuft im kombinierten Ansatz zunächst ebenso ab wie im nutzerbasierten Ansatz. Findet ein Knoten jedoch in seiner Nachbarschaft keine Bewertungen für ein Objekt, so führt er die Operation *suchen(Schlüssel(Objekt))* aus und erhält die Schlüssel der Nutzer, die das Objekt bewertet haben. Der verursachte Mehraufwand betrifft also lediglich Fälle, in denen sonst keine Empfehlung erzeugt werden könnte, und er beschränkt sich in diesen Fällen auf $(\eta + 1)$ Suchanfragen.

[5] Bei einem Standardfehler von 0,07.

Insgesamt lässt sich dieser Ansatz als Mittelweg zwischen nutzer- und objektbasiertem Ansatz einordnen. Im theoretisch ungünstigsten Fall ergibt sich für das Einfügen einer Bewertung und das Berechnen einer Empfehlung der gleiche Aufwand wie in der Summe aus objektbasiertem und rein nutzerbasiertem Ansatz: Der Aufwand des nutzerbasierten Ansatzes fällt immer an. Kann ein Knoten dann beim Einfügen einer Bewertung das eingefügte Objekt bei keinem seiner eigenen taste buddies finden, so kommt der Aufwand für den objektbasierten Ansatz noch hinzu. In den simulativ betrachteten Szenarien jedoch liegt der Aufwand für das Einfügen einer Bewertung nur geringfügig über dem rein nutzerbasierten Ansatz.

6.3 Datenschutz

Die drei Ansätze unterscheiden sich, neben ihrer Skalierbarkeit und der erzielbaren Empfehlungsqualität, auch in ihren Eigenschaften bezüglich des Datenschutzes:

- Der *objektbasierte* Ansatz erlaubt Zugriff auf die Information, welche Nutzer ein Objekt bewertet haben. Ist diese Information sensibel – beispielsweise im Fall von Medikamenten –, so bedeutet dies aus Sicht des Datenschutzes zunächst ein Problem. Diesem Problem kann jedoch durch die Verwendung von Pseudonymen begegnet werden. Ein Abruf aller Bewertungen eines Nutzers ist beim objektbasierten Ansatz nur durch eine „Brute-force"-Methode möglich, also durch Abrufen von Informationen über alle vorhandenen Objekte.

- Der *nutzerbasierte Ansatz* erlaubt Zugriff auf vollständige Nutzerprofile. Selbst die Verwendung von Pseudonymen birgt hier noch die Gefahr, dass die Identität eines Nutzers – beispielsweise durch das Vorhandensein einer ungewöhnlichen Objektkombination – aufgedeckt werden kann. Der nutzerbasierte Ansatz ist daher aus Sicht des Datenschutzes deutlich problematischer als der objektbasierte Ansatz. Jedoch ist die gezielte Suche nach einem Nutzer bei der Verwendung von Pseudonymen nicht möglich. Durch Anwendung des in Abschnitt 5.5.2.3 beschriebenen Verfahrens kann zudem ein Absuchen einer sehr hohen Anzahl von Nutzerprofilen deutlich erschwert werden. Schließlich kann der nutzerbasierte Ansatz auch mit dem in Abschnitt 4.4 beschriebenen Verfahren kombiniert werden; diese Möglichkeit wurde allerdings nicht evaluiert.

- Der *kombinierte* Ansatz hat grundsätzlich die gleichen Eigenschaften wie der nutzerbasierte Ansatz. Allerdings ermöglichen die Verweise für selten bewertete Objekte das Auffinden von Nutzern, die diese Objekte bewertet haben. Indem für jedes Objekt nur η Verweise gespeichert werden, wird dieses Risiko auf wenige Nutzer beschränkt. Ein

bösartiger speichernder Knoten – der allerdings das Objekt, für das er zuständig ist, nicht frei wählen kann – könnte allerdings eine Liste mit mehr als η Knoten speichern.

In allen Fällen können allerdings die juristischen Anforderungen an den Systemdatenschutz, wie sie in Abschnitt 3.6.7 dargestellt sind, erfüllt werden. Neben Unterrichtungs- und Einwilligungserfordernissen beinhalten diese

- die Ermöglichung der Beendigung der Nutzung des Dienstes (§ 13 Abs. 4 Nr. 1 TMG). Ein Knoten kann das Peer-to-Peer-System jederzeit verlassen.
- die Sicherstellung der sofortigen Löschung oder Sperrung der angefallenen Nutzungsdaten nach Zugriff (§ 13 Abs. 4 Nr. 2 TMG). Nutzungsdaten im Sinne des TMG werden nicht über den Prozess der Empfehlungserzeugung oder Bewertungsabgabe hinaus benötigt.
- den Schutz der Benutzung des Dienstes vor der Kenntnisnahme Dritter (§ 13 Abs. 4 Nr. 3 TMG). Dies betrifft (wie in Abschnitt 3.6.7 erörtert) nicht die Knoten, die an der Diensterbringung mitwirken, und ist somit ebenfalls sichergestellt.
- die Ermöglichung der getrennten Verwendung von Daten über die Nutzung verschiedener Telemedien (§ 13 Abs. 4 Nr. 4 TMG). Andere Telemedien als das beschriebene Empfehlungssystem werden an dieser Stelle nicht betrachtet.
- die lediglich pseudonyme Erstellung von Nutzungsprofilen, die nicht mit Daten über die Pseudonymträger zusammengeführt werden dürfen (§ 13 Abs. 4 Nr. 6 TMG). Nutzungsprofile, die die Nutzung des Empfehlungssystems selbst betreffen – nur diese werden durch die Regelung erfasst – werden nicht erstellt.
- die Anzeige der Weitervermittlung zu einem anderen Diensteanbieter (§ 13 Abs. 5 TMG). Eine solche Weitervermittlung findet nicht statt.
- die Ermöglichung anonymer oder pseudonymer Nutzung des Dienstes (§ 13 Abs. 6 TMG). Die Verwendung von Pseudonymen ist in allen Anwendungsfällen vorgesehen.
- die Pflicht zur Auskunftserteilung über gespeicherte, personenbezogene oder pseudonyme Daten (§ 13 Abs. 7 TMG). Diese Pflicht betrifft jeden einzelnen Knoten. Da Bewertungsdokumente signiert werden, ist es unproblematisch, zu verifizieren, ob ein Nutzer, der eine Auskunft begehrt, tatsächlich Träger des entsprechenden Pseudonyms ist. In diesem Fall kann die Auskunft erteilt werden.
- die Pflicht zur Anbieterkennzeichnung (§ 5 TMG). Diese ist in der Regel für Nutzer des Empfehlungssystems nicht einschlägig (vgl. Abschnitt 3.6.7.10) und im dargestellten Entwurf nicht vorgesehen; sie kann aber unproblematisch ergänzt werden.

6.4 Sicherheit

Die folgenden Sicherheitsanforderungen wurden für die Anwendung identifiziert:

- *Die Verfügbarkeit des Systems ist zu gewährleisten; ein Denial-of-Service-Angriff (DoS-Angriff) soll erschwert werden.* Denial-of-Service-Angriffe auf das Overlay-Netz sollen an dieser Stelle nicht betrachtet werden; diese sind für das Chord-Protokoll bereits in [SiMo02] diskutiert. Denkbar ist ein Denial-of-Service-Angriff aber auch auf Ebene der Anwendung. So könnte ein Angreifer das System durch massenweises Einfügen oder Abrufen von Dokumenten überlasten. Dies wird jedoch durch das in Abschnitt 5.5.2.3 beschriebene Verfahren verhindert. Der speichernde Knoten muss zunächst eine Operation – die Berechnung des Funktionswerts einer kryptographischen Hashfunktion – vornehmen, die jedoch nur geringen Rechenaufwand erfordert. Nur, wenn das Ergebnis dieser Berechnung die Bedingung aus Abschnitt 5.5.2.3 erfüllt, entsteht für den speichernden Knoten weiterer Rechen- und Speicheraufwand. In diesem Fall hat jedoch der Angreifer vorher einen höheren Aufwand – im Mittel muss er 2^{k-1} Funktionswerte der Hashfunktion errechnen, um die Bedingung – einen auf k Nullen endenden Funktionswert – zu erfüllen.

- *Das Zurückziehen einzelner Bewertungsdokumente – außer durch den Autor – ist zu verhindern.* Allen betrachteten Ansätzen ist gemein, dass Bewertungsdokumente mit einer Identität versehen und signiert werden. Erhält ein Knoten eine Anforderung, ein gespeichertes Dokument zu löschen, so tut er dies nur, wenn die Löschanforderung mit dem gleichen privaten Schlüssel signiert wurde wie das Dokument selbst.

  Abbildung 6.14 zeigt die Angriffe, die eine Löschung von Dokumenten aus dem System ermöglichen. Die Abbildung stellt einen Angriffsbaum [Schn99] dar, dessen Wurzel das eigentliche Angriffsziel ist. Die Bedingung eines Knotens ist erfüllt, sobald die Bedingung eines Tochterknotens erfüllt ist.

  - Der Angreifer könnte alle Knoten, die das Dokument speichern, unter seine Kontrolle bringen. Dies wäre möglich, wenn er
    - unter hinreichend vielen verschiedenen Identitäten im System auftreten könnte, so dass jeder der für ein Dokument zuständigen Knoten durch den Angreifer kontrolliert wird (Sybil-Angriff). Dieser Angriff wird durch die implementierten Verfahren nicht verhindert, sondern ist durch das darunter liegende Overlay-Netz zu lösen. Die Annahmen, die getroffen werden müssen, damit der Angriff nicht möglich wird, sind in Abschnitt 5.4 beschrieben. Abschnitt 2.1.7 zeigt, wie diese Annahmen durch in der Literatur beschriebene Ansätze erfüllt werden können.
    - eine gezielte Auswahl der zuständigen Knoten erreichen könnte. Auch dies

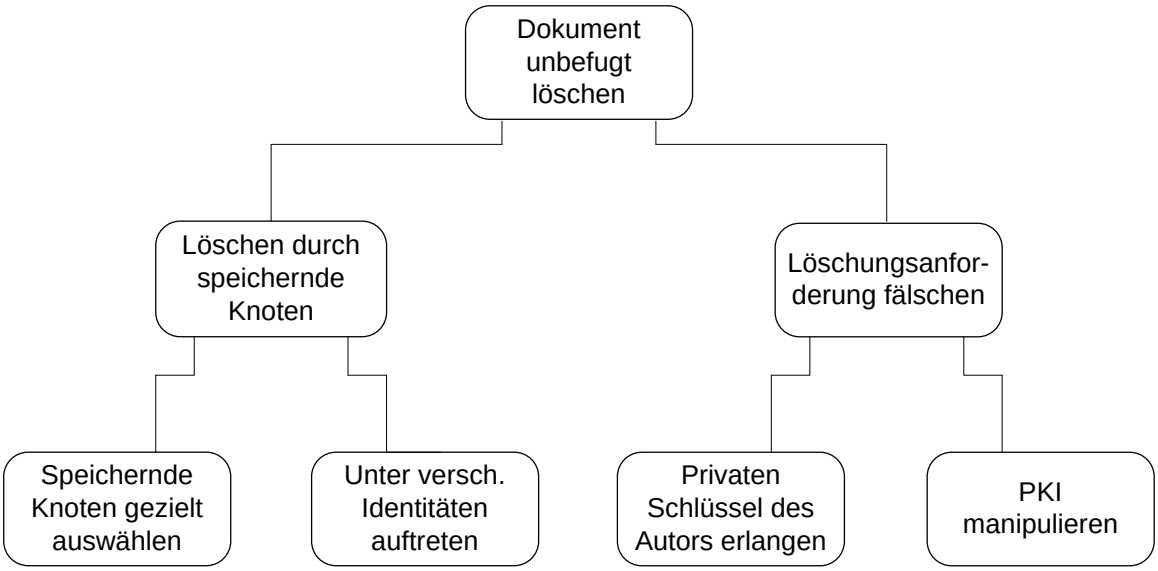

Abbildung 6.14: Angriffsbaum für die Löschung von Dokumenten

hängt vom zugrunde liegenden Overlay-Netz ab; Chord [SMKK+01] ermöglicht eine solche gezielte Auswahl nicht.

- Der Angreifer könnte eine Löschungsanforderung des Autors fälschen. Dazu müsste er
  - den privaten Schlüssel des Autors erlangen. Das entworfene System sieht keine Preisgabe privater Schlüssel vor; eine Weitergabe privater Schlüssel auf Kanälen außerhalb des Systems kann jedoch nicht verhindert werden.
  - die verwendete PKI manipulieren. Angriffe auf die PKI werden an dieser Stelle jedoch nicht betrachtet.

  Ähnliche Folgen wie eine Löschung hätte es auch, wenn der Angreifer das Auffinden der speichernden Knoten verhindern könnte. Dies wäre ein Spezialfall eines DoS-Angriffs auf das Overlay-Netz. Der Angriff führt zwar nicht zur Löschung der Dokumente, aber zu deren Unterdrückung und hat somit die gleichen Folgen. Auch dieser Angriff wird bereits in [SiMo02] betrachtet.
- *Die Integrität und Authentizität der gespeicherten Bewertungsdokumente ist zu gewährleisten.* Alle Dokumente sind signiert. In den grundlegenden Verfahren, die in Kapitel 5 betrachtet wurden, ergeben sich als Angriffsmöglichkeiten wiederum lediglich die Manipulation der verwendeten PKI und die Erlangung des privaten Schlüssels des Autors eines Dokuments.
- *Das massenweise Einspeisen von Bewertungsdokumenten soll erschwert werden.* Dieses Ziel wird mit dem gleichen Mechanismus erreicht, der auch DoS-Angriffe

verhindern soll. Zusätzlich ist im objektbasierten Ansatz als Erweiterung vorgesehen, die Anzahl an abgegebenen Bewertungen eines Knotens bei der Verwendung dieser Bewertungen zu berücksichtigen. Dies verringert den Anreiz, sehr viele Bewertungen abzugeben.

– *Der Einsatz eines Reputationssystems soll möglich sein.* Voraussetzung für den Einsatz eines Reputationssystems ist die Existenz von Identitäten, an die die Reputation angeknüpft werden kann. Grundsätzlich werden die Identitäten auf der Overlay-Ebene verwaltet und somit durch das vorgeschlagene Empfehlungssystem nicht beeinflusst. Die Ausnahme – die Erweiterung zum Identitätsschutz – wird in Abschnitt 6.4.2 diskutiert.

6.4.1 Angriffe auf Empfehlungsalgorithmen

Ein Angreifer könnte auch einen Anreiz haben, erzeugte Empfehlungen zu beeinflussen.

Im *nutzerbasierten* und somit auch im kombinierten Ansatz werden durch die Übermittlung vollständiger Bewertungsprofile die Angriffsmöglichkeiten eingeschränkt: Will ein böswilliger Knoten erreichen, dass ein bestimmtes Objekt anderen Knoten empfohlen wird, so ist sein Ziel, diesen anderen Knoten möglichst ähnlich zu erscheinen, um in ihrer Empfehlungsberechnung berücksichtigt zu werden. Im vorliegenden Ansatz reicht es aber nicht aus, Manipulationen auf Ebene einzelner Objektbewertungen vorzunehmen; die Ähnlichkeit wird auf Basis eines ganzen Profils errechnet. Kann ein Knoten nur die Ähnlichkeit zu wenigen anderen Knoten vortäuschen und gelangt nur in deren taste-buddy-Listen, so ergibt sich kein Problem, da er – auf das Gesamtsystem bezogen – nur geringen Schaden verursachen kann. Sichergestellt werden muss indes, dass ein Knoten nicht gegenüber unterschiedlichen Netzteilnehmern auch unterschiedliche Profile verwendet.

Das Ziel wird dadurch erreicht, dass den Knoten die Kontrolle darüber entzogen wird, wie sie sich bestimmten anderen Knoten gegenüber darstellen. Dies wird durch die Verwendung von Chord als Overlay-Netz und die darauf aufbauende Datenspeicherungsschicht garantiert. Will ein Knoten auf das Nutzerprofil eines anderen Knotens zugreifen, so kontaktiert er nicht den Knoten, dem das Nutzerprofil zuzuordnen ist, sondern die entsprechenden speichernden Knoten. Diese können durch den jeweiligen Nutzer nicht ausgewählt werden. Um zu gewährleisten, dass ein Knoten seine Chord-ID und somit die Schlüssel, für die er zuständig ist, nicht frei auswählen kann, muss auf der Overlay-Ebene allerdings ein geeignetes Verfahren zur Wahl der Identität herangezogen werden. Entsprechende Ansätze sind in [DiHa06], [Wall03] und [FrRe01] diskutiert.

Auch im *objektbasierten* Ansatz kann ein Knoten nicht kontrollieren, wie er sich einzelnen anderen Knoten gegenüber darstellt. Problematisch ist allerdings, dass keine vollständigen Profile vorliegen – andere Knoten bemerken also zunächst nicht, welche Bewertungen ein Knoten abgegeben hat. Zum einen wird ein massenhaftes Einfügen aber durch die Verwen-

dung von Crypto-Puzzles (Abschnitt 5.5.2.3) verhindert. Zum anderen werden Bewertungen von Knoten, die bereits viele Bewertungen abgegeben haben, geringer gewichtet – als Gewichtungsfunktion wird dabei $w_i(L_i) = \frac{1}{\log |L_i|}$ verwendet (vgl. Abschnitt 5.18). Schließlich muss, wie in Abschnitt 5.5 besprochen, dafür Sorge getragen werden, dass der Hersteller eines Objekts (Produkts) nicht durch die Wahl des entsprechenden Identifikators eine freie Platzierung in der Chord-DHT erreichen kann.

6.4.2 Erweiterung zum Schutz von Identitäten

Abschnitt 5.8.4.4 beschreibt eine Erweiterung, die verwendet werden kann, um die Identitäten von Teilnehmern bei der Aktualisierung einer Objekttabelle zu schützen. Die folgenden Angriffe auf diese Erweiterung sind nun denkbar:

- Der einfügende Knoten könnte versuchen, die Verknüpfung seiner Reputation mit der eingefügten Bewertung zu verhindern. Dazu müsste er
  - den Knoten H^k, der die Identität des einfügenden Knotens mit dem verwendeten Pseudonym verknüpft, auswählen bzw. durch einen kollaborierenden Knoten ersetzen können. Die Wahl dieses Knotens ist aber durch eine Funktion bestimmt, in die sowohl die Objekt-ID als auch die ID des bewertenden Knotens einfließt. Beide können durch den bewertenden Knoten nicht frei gewählt werden.
  - den lokal eindeutigen Bezeichner, den H^k ihm zusendet, manipulieren können. Dieser ist jedoch von H^k signiert. Die Veränderung des Bezeichners würde also eine Manipulation der PKI oder das Erlangen des privaten Schlüssels von H^k erfordern.
- Andere Knoten könnten versuchen, die Identität des Bewertenden in Erfahrung zu bringen:
  - Der Knoten H^k kennt die Identität des Bewertenden. Er hat jedoch keine Möglichkeit, eine Bewertung des Bewertenden abzurufen: Der Abruf von Bewertungen anhand der Identitäs des Bewertenden ist aber nicht möglich.
  - Der Knoten, der die Bewertung speichert, kennt das von H^k erzeugte Pseudonym. Nun könnte er
    - mit H^k zusammenarbeiten. In diesem Fall ist die Identität des Bewertenden offengelegt. Dies betrifft jedoch lediglich eine einzelne Bewertung; für andere Objekte sind wiederum andere Knoten H^k zuständig. Das heißt, dass ein einzelner bösartiger Knoten kein vollständiges Nutzerprofil aufdecken kann.
    - anhand der Identität von H^k versuchen, auf die Identität des bewertenden Knotens N zurückzuschließen. Die Identität von H^k ergibt sich daraus, dass

er zuständig ist für den Schlüssel (Erste $\frac{t}{2}$ bits von hash(Identifikator des neu bewerteten Objekts k) | Erste $\frac{t}{2}$ bits von hash(N)). Der Knoten, der die Bewertung speichert, kennt nun sowohl k als auch H^k. Um dadurch auf die Identität von N zurückschließen zu können, müsste er jedoch für alle denkbaren Identitäten von N den genannten Hashwert berechnen und prüfen, ob H^k für diesen zuständig ist.

– Dritte Knoten können die Objekttabelle vom speichernden Knoten abrufen; ihnen steht dann grundsätzlich die gleiche Information wie diesem speichernden Knoten zur Verfügung, und sie könnten sie in der gleichen Weise für Angriffe nutzen. Wie gezeigt, reiche diese Information aber nicht aus, um die Identität eines bewertenden Knotens offenzulegen.

Als Fazit lässt sich festhalten, dass das Offenlegen der Identität eines bewertenden Knotens nicht theoretisch unmöglich ist, da Brute-Force-Angriffe denkbar sind. Wohl aber wurde das Ziel erreicht, Angriffe so sehr zu erschweren, dass sie Kosten verursachen, die diese Angriffe wirtschaftlich unattraktiv machen können.

6.5 Fazit

Im vorliegenden Kapitel wurden die in Kapitel 5 vorgestellten Verfahren evaluiert. Dabei hat sich gezeigt, dass der objektbasierte Ansatz die beste Empfehlungsqualität liefert, allerdings der Kommunikationsaufwand für das Einfügen einer Bewertung mit der Anzahl von Objekten in den Nutzerprofilen der Nutzer steigt. Durch die vorgenommenen Optimierungen konnte im Vergleich zu in der Literatur betrachteten Ansätzen der Gesamtaufwand allerdings reduziert werden – bei konstanter Größe des Overlay-Netzes von $O(u^2)$ auf $O(u^{1,45})$.

Im Gegensatz dazu ist dieser Kommunikationsaufwand beim nutzerbasierten Ansatz konstant, doch kann hier in der Grundversion nur eine geringe Anzahl von Empfehlungen erzeugt werden.

Als Mittelweg hat der kombinierte Ansatz einen nur geringfügig höheren Aufwand für das Einfügen von Bewertungen und das Erzeugen von Empfehlungen als der nutzerbasierte Ansatz; die Anzahl erzeugter Empfehlungen kann dabei jedoch nahezu auf das Niveau des objektbasierten Ansatzes gesteigert werden.

Schließlich wurden auch Erweiterungen der Basisprotokolle zum Erreichen von Sicherheit und Datenschutz betrachtet.

In Kapitel 7 folgt nun eine juristische Betrachtung mit Schwerpunkt auf möglichen Fortentwicklungen des Rechts mit Bezug auf Peer-to-Peer-Systeme.

Kapitel 7

Vorschläge für eine Fortentwicklung des Rechts

Dieses Kapitel betrachtet Probleme der derzeitigen juristischen Einordnung von Peer-to-Peer-Netzen und macht Vorschläge für eine Fortentwicklung des Rechts. Dies betrifft zum einen sprachliche Fehlleistungen des Gesetzgebers, also Fälle, in denen die gesetzliche Regelung sprachlich vom gewollten Ziel abweicht; zum anderen werden teleologische Fehlleistungen betrachtet, also Regelungen, die inhaltlich nicht geeignet sind, das vom Gesetzgeber gewünschte Ziel zu erreichen.[1] Schließlich werden Wege zu einer sachgerechteren juristischen Behandlung der Peer-to-Peer-Netze untersucht.

7.1 Gesetzgeberische Fehlleistungen

Die Analyse soll zunächst von denjenigen in der Analyse aufgetretenen Problemen ausgehen, die als Fehlleistungen des Gesetzgebers eingeordnet werden können.

7.1.1 Sprachliche Fehlleistungen

Als sprachliche gesetzgeberische Fehlleistung ist wohl eindeutig die Formulierung der §§ 14, 15 TMG zu bezeichnen, die jeweils „nur“ die Erhebung und Verwendung von Bestandsdaten (§ 14) bzw. Nutzungsdaten (§ 15) durch einen Diensteanbieter erlauben; auf diese Weise wird sprachlich aus der Erlaubnis- zugleich eine Verbotsnorm gemacht. Der Gesetzgeber hat dies offensichtlich nicht beabsichtigt; Hintergrund der Einfügung des Wortes „nur“ in beide Vorschriften war der (systemwidrige) Versuch, eine Bußgeldvorschrift an die entsprechenden Vorschriften zu knüpfen. Die so veränderten Normen wurden später unverändert in das TMG übernommen (vgl. Abschnitt 3.6.2.1, S. 60). Eine Korrektur ist im Wege der (teleologischen) Auslegung möglich – dennoch ist dem Gesetzgeber anzuraten, mit der Bußgeldvorschrift des § 16 Abs. 2 Nr. 5 nicht an die §§ 14 Abs. 1,15 Abs. 1 anzuknüpfen, die ihrem Wesen nach Erlaubnisnormen sind. Da ohnehin keine Differenzierung

[1] Zu den Begriffen der sprachlichen und inhaltlichen Fehlleistungen siehe [Rüßm90, S. 43].

beispielsweise der Bußgeldhöhe nach Art der erhobenen oder verwendeten Daten stattfindet, wäre eine Anknüpfung an die Verbotsnorm des § 12 Abs. 1 TMG bereits zielführend.

Auf eine weitere sprachliche Fehlleistung weist Bizer [Bize07] hin: Telekommunikationsgestützte Dienste nach § 3 Nr. 25 TKG werden vom Anwendungsbereich des TMG ausgenommen (§ 1 Abs. 1 Satz 1 TMG). Folgte man dem Wortlaut der Definition („Dienste, die keinen räumlich und zeitlich trennbaren Leistungsfluss auslösen, sondern bei denen die Inhaltsleistung noch während der Telekommunikationsverbindung erfüllt wird"), so wären die meisten Peer-to-Peer-Anwendungen und außer ihnen ein großer Teil der weiteren im Internet angebotenen Dienste keine Telemedien – ein Ergebnis, dass der Gesetzgeber nicht beabsichtigte.[2]

7.1.2 Teleologische Fehlleistung

Stellt sich ein Rechtsproblem im Zusammenhang mit Peer-to-Peer-Systemen, so ist der erste Schritt der Betrachtung die Fragestellung, wie sich das Problem in das Schichtenmodell der Gesetzgebung einordnen lässt: Ist das Telekommunikationsrecht zu betrachten, ist es das Telemedienrecht oder die allgemeinen Gesetze – oder eine Kombination von mehreren dieser Möglichkeiten? Diskutiert werden soll an dieser Stelle indes nur ein Teilbereich dieser Problematik, nämlich der Bereich des Datenschutzes.

Vergleichsweise unproblematisch ist die Entscheidung zwischen den „obersten" Schichten (also zwischen Telemedienrecht und den allgemeinen Gesetzen), wie dies für das Datenschutzrecht bereits in Abschnitt 3.6.1 (S. 58) gezeigt wurde: Sobald die Daten einen Bezug zum betrachteten Telemedium haben, sind die Datenschutzbestimmungen des TMG anwendbar – dies gilt nur dann nicht, wenn das Telemedium lediglich der Übermittlung von Daten dient, die auch ohne das Vorliegen eines Telemediums denkbar wären. Auch jenseits des Datenschutzrechts ist das Telemediengesetz lediglich relevant, wenn für einen elektronischen Informations- oder Kommunikationsdienst spezifische Fragestellungen zu betrachten sind. Schließlich war auch schon innerhalb dieser Schicht nach altem Recht die Abgrenzung zwischen Telediensten und Mediendiensten zwar nicht immer eindeutig[3], kam aber in der Praxis aufgrund meist übereinstimmender Regelungen in TDG und MdStV nur selten zum Tragen.

Sehr viel problematischer stellt sich jedoch die Abgrenzung zwischen Telekommunikation und Telemedien dar – eine Problematik, die der Gesetzgeber so nicht erkannt hat (vgl. dazu Abschnitt 3.6.2.1), so dass von einer teleologischen Fehlleistung gesprochen werden kann. Bereits nach altem Recht war die Trennlinie schwierig zu ziehen: Auch, wenn in der Literatur teilweise die Ansicht vertreten wird, diese Abgrenzung lasse sich anhand

[2] So auch Bizer [Bize07]; auch die Begründung des TMG-Entwurfs [Bund06a, S. 11] spricht davon, dass die geltenden Vorschriften aus TDG und MdStV weitgehend unverändert bleiben sollen – mit einer derart drastischen Beschneidung des Geltungsbereichs wäre dies indes nicht mehr möglich.

[3] So [Wald98, S. 124], [Goun97, S. 2994]; vgl. auch Abschnitt 3.5.1.1.

der Schichtengrenzen des ISO/OSI-Schichtenmodells (bzw. des für die Praxis relevanteren TCP/IP-Schichtenmodells) vornehmen[4], ist dies tatsächlich nicht möglich. Dies zeigt sich gerade bei Overlay-Netzen – wie den im Rahmen dieser Arbeit betrachteten Peer-to-Peer-Systemen –, da hier innerhalb der Anwendungsschicht (aus Sicht des Underlays) wiederum Funktionalitäten implementiert sind, die sich eigentlich unteren Schichten des Schichtenmodells zuordnen lassen. Die Verabschiedung des Telemediengesetzes hat diesem Problem nur teilweise abgeholfen. So wurde klargestellt, dass es Dienste geben kann, die dem Regelungsregime sowohl des TKG als auch des TMG unterliegen: Dies ergibt sich daraus, dass § 1 Abs. 1 Satz 1 des Gesetzes aus dem Anwendungsbereich des TMG lediglich telekommunikationsgestützte Dienste (§ 3 Nr. 25 TKG) und Telekommunikationsdienste (§ 3 Nr. 24 TKG), „die *ganz* in der Übertragung von Signalen über Telekommunikationsnetze bestehen" ausschließt. Nicht ausgeschlossen sind jedoch solche Dienste, die lediglich *überwiegend* in der Übertragung von Signalen über Telekommunikationsnetze bestehen (und ebenfalls von der Definition eines Telekommunikationsdienstes in § 3 Nr. 24 TKG erfasst sind). Es handelt sich hierbei also sowohl um Telemedien als auch um Telekommunikationsdienste – eine Konstellation, die der Gesetzgeber auch so beabsichtigt hat.[5] Für diese Klasse von Telemedien ist allerdings wiederum ein Großteil der Datenschutzregelungen des TMG ausgeschlossen (§ 11 Abs. 3 TMG).

Die Abgrenzung zwischen Telekommunikationsdiensten und Telemedien indes hat sich durch die neue Regelung zwar verschoben, aber nicht unbedingt vereinfacht. So sind nach neuer Rechtslage drei Abstufungen zu unterscheiden: Reine Telekommunikationsdienste, reine Telemedien und die Kombination aus beiden. Die Entscheidung, wann welcher dieser Fälle vorliegt, dürfte sich in der Praxis nicht immer einfach gestalten.

Wenig gelungen ist die Abgrenzung zwischen TMG und TKG auch in anderer Hinsicht. So gilt das Telemediengesetz nach seinem § 1 für Telemedien; der Begriff des Diensteanbieters geht aber, indem er Personen beinhaltet, die lediglich „fremde Telemedien zur Nutzung [bereithalten] oder den Zugang zur Nutzung [vermitteln]" (§ 2 Nr. 1 TMG), über diesen Anwendungsbereich ebenso hinaus wie die Verantwortlichkeitsregelung des § 8 Nr. 1 TMG, die Zugangsprovider von der Verantwortlichkeit für übermittelte Inhalte freistellt. Konsequent wäre eine Beschränkung des Diensteanbieterbegriffs auf Anbieter von Telemedien sowie eine Überführung der Regelung des § 8 Nr. 1 TMG in das Telekommunikationsgesetz (zu dieser schon im Verhältnis zwischen TDG und TKG bestehenden Problematik bereits Stadler [Stad05, Rn. 33]).

[4] So beispielsweise in [Raab03, S. 136]; vgl. dazu auch Abschnitt 3.2.2.1
[5] Siehe hierzu die Gesetzesbegründung [Bund06a, S. 13].

7.2 Gliederungsansatz

Die Lösung der aufgezeigten Problematik – der unklaren Abgrenzung zwischen den „Schichten“ der Gesetzgebung – kann durch die Wahl eines anderen Gliederungsansatzes gefunden werden. Dies gilt insbesondere für den Bereich des Datenschutzes.

Bereits im Jahr 2000 wurde – im Zusammenhang mit dem Erlass der Telekommunika tions-Datenschutzverordnung – eine Zusammenlegung der Datenschutzregelungen für Telekommunikations- und für Teledienste gefordert [KoNe00, S. 425]; diese Forderung wurde während des Gesetzgebungsverfahrens zur Novellierung des TKG dann auch durch den Deutschen Bundestag erneut erhoben[6]. Tatsächlich umgesetzt wurde sie bislang nicht.

7.2.1 Schutz nach Risiko

Ein erster Schritt kann darin bestehen, die Datenschutzregelungen nach dem Ausmaß des Risikos zu gliedern, dem der Nutzer ausgesetzt ist. Das Risiko selbst lässt sich schwerlich messen, weshalb stellvertretende Variablen herangezogen werden müssen, an die die Gesetzgebung anknüpfen kann. Hierbei können folgende Grundsätze herangezogen werden:

- *Schutz der Individualkommunikation zwischen Menschen* Ein besonders hohes Schutzniveau ist bei der Individualkommunikation zwischen Menschen erforderlich – unabhängig davon, wie diese technisch realisiert wird. Dies gilt sowohl für die Inhalte dieser Kommunikation als auch für ihre näheren Umstände. Grund für dieses hohe Schutzbedürfnis ist der oft persönliche Charakter des zwischen Menschen ausgetauschten gesprochenen oder geschriebenen Worts.

- *Schutz der Privatsphäre in der Mensch-Computer-Kommunikation* Die Kommunikation zwischen Mensch und Computer – beispielsweise, weil der Mensch einen im Computersystem realisierten Dienst in Anspruch nehmen möchte – ist *an sich* unproblematisch. Ein Problem entsteht dann, wenn der Kommunikationsvorgang Rückschlüsse auf personenbezogene Daten einzelner Nutzer erlaubt. Dies kann bereits die Tatsache sein, dass ein Nutzer mit einem bestimmten Rechner (oder einem von diesem Rechner angebotenen Dienst) kommuniziert hat, aber auch inhaltliche Aspekte oder den Kontext (Aufenthaltsort des Nutzers, Zeitpunkt der Kommunikation) beinhalten. Auch weitere hinzukommende Umstände können hier von Bedeutung sein – beispielsweise die Frage, ob und inwieweit der Nutzer sich der Tatsache bewusst ist, dass personenbezogene Daten durch das Computersystem erhoben und verarbeitet werden. Irrelevant ist jedoch, auf welcher Schicht des ISO/OSI-Schichtenmodells die entsprechenden Daten anfallen. Wird der Nutzer durch seine IP-Adresse identifiziert oder hat er sich auf einer besuchten Website registriert und wird durch Verwendung eines

[6] Einem entsprechenden Antrag [TaGV02] stimmte der Bundestag im Juli 2002 zu [Bund02, S. 25195]

Cookies wiedererkannt, ist die Konsequenz – die Identifikation einer Person oder doch zumindest eines kleinen Personenkreises – die gleiche. Dass hier überhaupt ein Spezialfall der Gefährdung informationeller Selbstbestimmung vorliegt, der über die allgemeine Datenschutzproblematik hinausgeht, liegt im wesentlichen an der mangelnden Kontrolle, die der Nutzer bei einer Vielzahl von Telekommunikationsvorgängen mit Rechnern unbekannter Betreiber ausüben kann.

- *Umfang anfallender Daten* Als wesentliches Kriterium für die Regulierung kann auch der Umfang dienen, in dem Daten erhoben werden können. Dies kann auch als eine Motivation für die Unterscheidung zwischen Diensteanbieter und Nutzer im Telemediengesetz gesehen werden: Zwar kann der Nutzer auch Daten über den Anbieter erlangen, die durchaus Personenbezug haben können. Doch kann – insbesondere bei Abruf- und Verteildiensten, die dem Gesetzgeber offenbar bei Verabschiedung des TMG vor Augen standen – ein einzelner Anbieter in der Regel Daten über sehr viele Nutzer sammeln. Neben der Differenzierung zwischen Anbieter und Nutzer wären weitere Kriterien denkbar, die als Stellvertreter an den Umfang anfallender personenbezogener Daten (und somit das Risiko für den Betroffenen) anknüpfen, so beispielsweise der Umfang der angebotenen Dienstleistung oder deren Gewerblichkeit bzw. Geschäftsmäßigkeit.

- *Kontrolle* Ebenso kann das Ausmaß der Kontrolle, die der Betroffene über die Preisgabe seiner Daten hat, herangezogen werden. Auch aus dieser Perspektive lässt sich die Unterscheidung zwischen Diensteanbieter und Nutzer im Telemediengesetz erklären: Bei „klassischen“ Telemedien wie beispielsweise Websites ist für den Anbieter meist offensichtlich, welche (personenbezogenen) Daten er preisgibt; dem Nutzer ist dies nicht unbedingt klar. Werden Cookies übertragen oder speichert der Anbieter die IP-Adresse des Nutzers, bemerkt dieser das in aller Regel gar nicht.

Wie sich zeigt, werden unterschiedliche Risikopotentiale durch die Gesetzgebung also durchaus schon berücksichtigt. Das Aufkommen von Peer-to-Peer-Systemen stellt die Annahmen, die der Gesetzgebungssystematik zugrund liegen, jedoch in Frage. Insbesondere entspricht die Unterscheidung zwischen Anbietern und Nachfragern von Diensten nicht mehr der Sachlage. Dies liegt nicht nur an der jederzeit vorliegenden Möglichkeit zum Rollenwechsel. Auch sonst ist praktisch keine Asymmetrie mehr gegeben: Ein Nutzer interagiert ebenso mit vielen unterschiedlichen Anbietern wie ein Anbieter mit unterschiedlichen Nutzern. Selbst die Natur der preisgegebenen Daten muss sich nicht zwangsläufig zwischen beiden Rollen unterscheiden, und die Transparenz über erhobene und verarbeitete Daten ist für den menschlichen Benutzer unabhängig davon, ob er gerade als Anbieter oder als Nachfrager agiert.

Als Konsequenz kann festgehalten werden, dass die Unterscheidung zwischen allgemeinen Regelungen des Datenschutzes und besonderen Regelungen für den Fall von Telekommunikationsbeziehungen nach wie vor sinnvoll ist. Auch eine Unterscheidung nach Art der erhobenen oder genutzten Daten – ähnlich den Regelungen des BDSG – ist nach wie vor denkbar. Dies gilt ebenso für eine Differenzierung zwischen geschäftsmäßig handelnden und sonstigen Teilnehmern.

Die Abgrenzung zwischen den verschiedenen Schichten der Gesetzgebung alleine kann zwar als wichtiger Schritt zu einem praktikableren Umgang mit dem Recht der Peer-to-Peer-Systeme führen; es sind aber weitere Schritte vorstellbar, mit denen das Recht den Charakteristika solcher Netze noch besser gerecht werden könnte. Diese werden in den folgenden Abschnitten diskutiert.

7.3 Weitergehende Ansätze

7.3.1 Ausgangssituation

Dass eine zufriedenstellende Gesetzgebung, die dem Wesen von Peer-to-Peer-Systemen gerecht wird, nur schwer gefunden werden kann, liegt wesentlich in den technischen Eigenschaften solcher Netze und daraus resultierenden gegensätzlichen Interessen begründet. Solche Interessenkonflikte werden in den folgenden Abschnitten beschrieben.

7.3.1.1 Datenschutz vs. Verfolgbarkeit

Ein offensichtlicher Interessensgegensatz zeigt sich zwischen Datenschutz einerseits und der Verfolgbarkeit von Rechtsverstößen – sowohl aus strafrechtlicher Sicht als auch bei der Verfolgung zivilrechtlicher Ansprüche – andererseits. Werden technische Lösungen zum Datenschutz auf der Empfänger-, Kontroll-, Zweck- oder Inhaltsebene realisiert, so stehen Möglichkeiten offen, die Zurückverfolgbarkeit von Aktionen in besonderen Fällen (beispielsweise durch richterlichen Beschluss) zu ermöglichen, ohne das Datenschutzniveau im Allgemeinen dadurch zu reduzieren. Dies liegt daran, dass in diesen Ebenen der Personenbezug von Daten grundsätzlich erhalten bleibt; eingeschränkt wird lediglich der Umgang mit diesen Daten.

Der technische Datenschutz in Netzen wird in der Praxis jedoch überwiegend auf der Identitätsebene – beispielsweise mit Verfahren wie Onion Routing [ReSG98] – umgesetzt. Zwar ist es theoretisch denkbar, auch hier die Möglichkeit der nachträglichen Offenlegung einer Identität vorzusehen – im einfachsten Fall durch ausschließliche Verwendung von Pseudonymen, die durch einen vertrauenswürdigen Dritten vergeben werden und durch diesen gegebenenfalls offengelegt werden können. Es bestehen jedoch keine Anreize für

Netzteilnehmer, eine solche Lösung tatsächlich einzusetzen. Somit führt das Verstecken der Identität zur Nichtverfolgbarkeit der Netzteilnehmer.

7.3.1.2 Schutz der Nutzer vs. Einfachheit für Anbieter

Eine gerade für das Telekommunikationsrecht wichtige Interessenskonfliktlinie verläuft zwischen der Einfachheit, einen Dienst anzubieten, und dem Schutz des Nachfragenden. Durch Regulierung – beispielsweise im Bereich des Datenschutzes – kann der Dienstnutzer erheblich besser gestellt werden. Beispielsweise kann die Pflicht zur Anbieterkennzeichnung ihm die Verfolgung von Ansprüchen erleichtern; Regelungen zum Systemdatenschutz erschweren den Missbrauch personenbezogener Daten, und Unterrichtungspflichten sorgen dafür, dass dem Nutzer der Umfang der Verarbeitung seiner personenbezogenen Daten bewusst ist. Der Anbieter andererseits hat die Lasten dieser Bestimmungen zu tragen; dies kann nicht nur zu höheren Kosten führen, sondern auch den Schutz personenbezogener Daten *des Anbieters* beeinträchtigen. Diese Besserstellung des Nutzers als schwächerer Partei ist auch gerechtfertigt, hält man sich die angenommene Asymmetrie vor Augen, die sowohl dem TMG als auch dem TKG zugrunde liegt: Der Anbieter interagiert typischerweise mit einer Vielzahl von Nutzern; die Kosten, die durch Regulierung entstehen, fallen jedoch zum größten Teil nur einmalig an, sodass die pro Nutzer zu tragenden Kosten kaum mehr ins Gewicht fallen. Andererseits ist das Gefährdungspotential hoch, da gerade bei Kommunikationsvorgängen das informationelle Selbstbestimmungsrecht des Nutzers betroffen ist und insbesondere durch den Anbieter leicht verletzt werden kann. Die Besonderheit bei Peer-to-Peer-Systemen ist nun der weitgehende Wegfall dieser Asymmetrie, da jeder Teilnehmer des Netzes grundsätzlich beide Rollen übernehmen kann. Es stellt sich also die Frage, ob und inwieweit sich Verpflichtungen noch halten lassen, die aus der Motivation der Asymmetrie heraus begründet wurden.

7.3.2 Weitere Problemfelder

Auch unabhängig von diesen entgegengesetzten Interessen gilt es bei Peer-to-Peer-Systemen diverse Besonderheiten zu beachten:

- Ein Dienst wird oft durch das System als Ganzes angeboten; einzelne Knoten realisieren oft nur einen kleinen Teil des Dienstes. Einwirkungsmöglichkeiten durch gesetzgeberische Maßnahmen wiederum können nur einzelne Komponenten beeinflussen.
- Peer-to-Peer-Systeme erlauben das Entstehen von Gemeinschaften, aber auch völlige Anonymität zwischen den Teilnehmern.
- Durch ein Peer-to-Peer-System kann ein einzelnes Softwareprodukt, das ein bestimmtes Protokoll implementiert, eine erhebliche Nutz- oder Schadwirkung entfalten, die

erst durch die Vernetzung möglich wird und sehr viel schwieriger kontrollierbar ist als bei Vorhandensein zentraler Komponenten. Diese Wirkung entsteht durch das Verhalten der Nutzer, wird aber durch das Softwareprodukt ermöglicht.

In der Summe kann diese Entwicklung zu einer für das Recht äußerst problematischen Situation führen: Durch das Gesamtsystem können Wirkungen entstehen, die durch das Rechtssystem missbilligt werden. Jedoch können potentiell mehrere Gründe zusammenkommen, die dazu führen, dass diese unter Umständen nicht auf einzelne Teilnehmer zurückgeführt werden:

- Der durch einen einzelnen Teilnehmer geleistete Beitrag ist zu klein, um – zumindest nach derzeitiger Rechtslage – Sanktionen gegen diesen Teilnehmer zu rechtfertigen.
- Der durch einen einzelnen Teilnehmer verursachte Schaden ist zu klein, als dass sich die Durchsetzung von Schadensersatzforderungen gegen einzelne Teilnehmer lohnen würde.
- Das System ist hinreichend robust, um die unerwünschte Wirkung aufrecht zu erhalten, auch, wenn Unterlassungsansprüche gegen einzelne Teilnehmer durchgesetzt werden können.
- Die Teilnehmer sind anonym und können nicht ermittelt werden.

Zusammengenommen kann von einer Diffusion der Verantwortlichkeit gesprochen werden; wo vorher ein einzelner Anbieter zur Rechenschaft gezogen werden konnte, steht ein Geschädigter oder Betroffener nun einer Vielzahl möglicher Anspruchsgegner gegenüber. Im nächsten Abschnitt sollen Lösungsansätze für diese Problematik diskutiert werden.

7.4 Lösungsansätze

Lösungsansätze für die dargestellten Probleme könnten sich aus der Betrachtung bereits untersuchter Systeme ergeben, die zumindest teilweise mit Peer-to-Peer-Systemen vergleichbar sind.

7.4.1 Systeme mit vergleichbarer Interessenlage

Ein allgemeines Problem selbstorganisierender Systeme ergibt sich für das Recht bereits aus deren Definition. Ein selbstorganisierendes System zeigt emergentes Verhalten; das bedeutet, dass einfache Interaktionen zwischen Systemteilnehmern die Erfüllung auch komplexer Aufgaben durch das Gesamtsystem ermöglichen. Daraus folgt auch, dass es in einem solchen System in der Regel keine Instanz mehr gibt, die alleine für das Systemverhalten

verantwortlich zu machen wäre. Im Beispiel des Peer-to-Peer-Systems kann insbesondere der einzelne Nutzer keinen wesentlichen Einfluss auf die Funktionalität des Gesamtsystems ausüben.

7.4.1.1 Softwareagenten

Eine im Grundsatz ähnliche Situation findet sich im Bereich der Softwareagenten. Es handelt sich um Softwaresysteme, die in einer bestimmten Umgebung[7] *autonom* agieren können, um ein bestimmtes Entwurfsziel zu erreichen.[8] Autonom ist ein System, wenn es in einem gewissen Umfang sein eigenes Verhalten kontrollieren und ohne den Eingriff von Menschen oder anderen Systemen handeln kann [Weis99, S. 2]. Agenten als autonome Systeme erfüllen somit bereits einen Teil der Definition selbstorganisierender Systeme. Dennoch sind sie aus zivilrechtlicher Sicht nach derzeitiger Rechtslage nicht anders einzuordnen als andere Computerprogramme auch [Sorg05, S. 23–36]. Ein Agent ist selbst kein Rechtssubjekt, und durch ihn abgegebene Willenserklärungen sind aus rechtlicher Sicht Willenserklärungen des Nutzers, der seinen Agenten für deren Abgabe lediglich als Werkzeug einsetzt. Eine solche Lösung ist auch angemessen, denn sie weist das Risiko des Agenteneinsatzes demjenigen zu, der auch dadurch profitiert. Problematisch bei dieser Einordnung ist allein der Geschäftswille, also der Wille des Erklärenden, eine konkrete Rechtsfolge herbeizuführen (vgl. [Sorg05, S. 27 f.]). Auch wenn sich Agenten – beispielsweise durch den verstärkten Einsatz maschineller Lernverfahren – zukünftig noch autonomer handeln sollten, als dies bisher der Fall ist, bleibt es aber dabei, dass nur bei der Annahme eines vorhandenen Geschäftswillens auch in Bezug auf eine einzelne Willenserklärung eine angemessene Risikoverteilung erreicht wird [Sorg05, S. 119].

Wie lassen sich diese Ergebnisse nun auf selbstorganisierende Peer-to-Peer-Systeme übertragen? Zwar erscheint der Einsatz solcher Systeme zur autonomen Erstellung von Willenserklärungen eher unwahrscheinlich, doch lassen sich die Grundsätze der Betrachtung auch auf die Verantwortlichkeit in anderen Bereichen übertragen. Wiederum sollte das Risiko demjenigen zugewiesen werden, der sich für den Einsatz des Systems entscheidet und den entsprechenden Nutzen trägt.

Wesentlicher Unterschied zwischen (einzelnen) Agenten und den in dieser Arbeit betrachteten Peer-to-Peer-Systemen ist die Emergenz. Dies bedeutet, dass einfache Interaktionen zwischen den beteiligten Entitäten zu einem komplexen Systemverhalten führen können. Zudem ist zumindest derzeit nicht absehbar, dass Peer-to-Peer-Systeme – im Gegensatz zu einzelnen Agenten – zur Erstellung von Willenserklärungen eingesetzt werden.[9]

[7] Im weiteren Sinne; auch eine elektronische Handelsplattform kann eine solche Umgebung sein

[8] An *agent* is a computer system that is *situated* in some *environment*, and that is capable of *autonomous action* in this environment in order to meet its design objectives [Wool02, S. 15].

[9] Dies ist unabhängig davon, daß in Peer-to-Peer-basierten Marktplattformen Willenserklärungen auf einzelnen Knoten erzeugt und dann über das Peer-to-Peer-System übermittelt werden können.

Auf Peer-to-Peer-Systeme übertragen werden könnte aber die Idee, dass auch bei einer Abkopplung des Systemverhaltens von den Absichten des jeweiligen Nutzers dieser jedenfalls dann das Risiko zugewiesen bekommt, wenn er die Entscheidung für die Teilnahme am System getroffen hat und durch den Einsatz des Systems selbst profitiert. Das Problem bei dieser Einstufung ist – bezogen auf Peer-to-Peer-Systeme – jedoch, dass

- der einzelne Teilnehmer oft gar nicht absehen kann, wie sich das System verhalten wird; er hat zwar möglicherweise Kontrolle über eine Systemkomponente, keinesfalls aber über das Zusammenwirken vieler Komponenten.
- angesichts einer Vielzahl von Systemteilnehmern nur schwerlich entschieden werden kann, *welchem* dieser Teilnehmer ein Risiko zugewiesen werden soll.

Die Einwirkungsmöglichkeit des Rechts auf das Gesamtsystem ist begrenzt; das System ist weder Rechtssubjekt noch kann es durch ein einzelnes Rechtssubjekt ohne weiteres kontrolliert werden. Doch wie in den folgenden Abschnitten gezeigt wird, ist dieses Problem nicht neu.

7.4.1.2 Umweltrecht

In der Tat wird emergentes Verhalten auch für andere Systeme postuliert, mit denen das Recht sich befassen muss. Leydesdorff [Leyd93] beispielsweise diskutiert – ausgehend von einem Begriff der Selbstorganisation, der im Wesentlichen auf dem Vorhandensein emergenten Verhaltens beruht [Leyd93, S. 331] – die Frage, ob eine Gesellschaft als selbstorganisierendes System eingeordnet werden kann. Auch, wenn der Autor die Antwort auf diese Frage offenlässt, so lassen sich doch Parallelen ziehen, weil auch das Recht nur indirekt auf das Gesellschaftssystem als ganzes einwirken kann. Deutlicher wird dies noch beim System Umwelt: Haben beispielsweise die Emissionen einer einzelnen Entität (beispielsweise eines Unternehmens) noch keinen bezifferbaren Einfluss auf das System Umwelt als Ganzes, so stellt sich die Frage, wer für eventuelle Schäden haftbar gemacht werden soll. Für den Fall von Umweltbeeinträchtigungen, die keinem einzelnen Schädiger zugeordnet werden können – sogenannte Summationsschäden – , hat der Gesetzgeber allerdings keine Lösung gefunden. Das Umwelthaftungsgesetz (UmweltHG), das Haftungsregelungen für den Eingriff in die Umwelt vorsieht, lässt solche Schäden bewusst außen vor [Bund90, S. 16]. Lediglich, wenn sich eigene, konkrete Gefährdungsbeiträge abgrenzen lassen, tritt auch die Kausalitätsvermutung des § 6 Abs. 1 Satz 1 UmweltHG ein (Hager in [HaRe07, Rn. 37 zu § 6 UmweltHG]). Im Zusammenhang mit solchen Summationsschäden besteht auch keine Haftung der öffentlichen Hand [BGH88]. Andererseits ist auch eine gesamtschuldnerische Haftung von Beteiligten denkbar. Als Anspruchsgrundlage kommt § 830 Abs. 1 BGB in Frage, der eine gesamtschuldnerische Haftung im Fall der gemeinschaftlichen Begehung einer

unerlaubten Handlung sowie im Fall, dass bei mehreren Beteiligten sich der Schadensverursacher nicht ermitteln lässt, begründet. Voraussetzung dieser gesamtschuldnerischen Haftung ist allerdings die Gesamtkausalitätseignung – die Eignung jedes einzelnen Tatbeitrags, bereits alleine den Gesamtschaden zu verursachen (Wagner in [ReSR04, Rn. 50 zu § 830 BGB]).

Die Übertragung auf Systeme, welche sich dadurch auszeichnen, dass der Beitrag einzelner Teilnehmer zum Systemverhalten nicht identifiziert werden kann, ist nach derzeitiger Rechtslage nicht möglich. Sie wäre auch nicht sachgerecht, da sie ohne Verschulden, sogar ohne nachgewiesene Kausalität zu einem erheblichen Haftungsrisiko der Teilnehmer führen würde. In der amerikanischen Rechtsprechung wurden bei Verursachung eines Schadens durch mehrere Täter bereits Einzelne dieser Täter zum Ersatz des Gesamtschadens verurteilt, sofern sie den durch sie begründeten Schadensanteil nicht nachweisen konnten [Hage86, S. 1969]. Auch diese Konstruktion setzt allerdings voraus, dass wenigstens ein rechtswidrig handelnder und schadensverursachender Täter identifiziert werden kann.

Auch, wenn einiges für die Vergleichbarkeit von Problemen, wie sie beispielsweise im Umweltrecht auftreten, mit denen von Peer-to-Peer-Systemen spricht, so bringt die Abkopplung von der physischen Welt doch auch neue Herausforderungen mit sich:

- Der Nachweis eines Schadens, der Kausalität eines Ereignisses oder einer Handlung für diesen Schaden und der Identität des Schadensverursachenden ist durch das Fehlen physischer Spuren erschwert; der Einsatz von Anonymisierungstechniken verschärft dieses Problem.
- Im Einzelfall entstehende Schäden sind oft sehr klein, weshalb es oft wirtschaftlich nicht sinnvoll ist, den Ersatz dieser Schäden einzufordern. Die Summe zahlreicher kleiner Schäden kann jedoch wiederum wirtschaftlich bedeutsam sein.
- Lassen sich die genannten Probleme innerhalb eines Staates möglicherweise noch lösen, so führt die Internationalität des Internets und darauf aufbauender Netze zu einer erheblichen Verschärfung aufgrund unterschiedlicher Regelungsregimes in den einzelnen Staaten.
- Schließlich sind auch neue Regelungsbereiche betroffen, so insbesondere das Urheberrecht.

Deshalb sollen neben dem Umweltrecht noch weitere Regelungsbereiche betrachtet werden.

7.4.1.3 Personengesellschaften

Ein naheliegendes Rechtsgebiet, in dem sich vergleichbare Probleme ergeben, ist das Gesellschaftsrecht. Die Betrachtung in Abschnitt 3.5.2.2 (S. 54) hat zwar ergeben, dass nach ak-

tueller Rechtslage ein Peer-to-Peer-System nicht als Personen- oder Kapitalgesellschaft eingeordnet werden kann. Dennoch sollen exemplarisch Instrumente des Gesellschaftsrechts betrachtet und untersucht werden, inwiewiet sich diese auf Peer-to-Peer-Systeme übertragen lasen.

- *Haftungsregelungen* können dazu dienen, eine angemessene Risikoverteilung zwischen Gesellschaftern und Dritten zu erreichen. Das Problem bei diesem Ansatz ist bei einem selbstorganisierenden System jedoch die konkrete Ausgestaltung solcher Regelungen. Eine gesamtschuldnerische Haftung der Systemteilnehmer beispielsweise würde einem Verbot des Systems gleichkommen, da das so entstehende Risiko für den Einzelnen nicht tragbar wäre. Da alle Teilnehmer die gleichen Rollen einnehmen, ist zumindest in Peer-to-Peer-Systemen auch keine andere angemessene Haftungsregelung vorstellbar.
- Auch schafft das Gesellschaftsrecht mit juristischen Personen einen künstlichen Anknüpfungspunkt für Rechte und Pflichten; dies erleichtert den Umgang Dritter mit einer Gesellschaft. Hilfreich ist dies jedoch nur, wenn dieser Anknüpfungspunkt auch mit einem eigenen Vermögen ausgestattet wird, denn eventuell entstehende Ansprüche müssen auch befriedigt werden können. Eine solche Konstruktion ist für Peer-to-Peer-Systeme allenfalls theoretisch denkbar; praktisch würde sie zu hohen Transaktionskosten führen, und es ist unklar, wie die Ausstattung mit einem eigenen Vermögen gestaltet werden könnte. Denkbar wäre die Erhebung einer Gebühr von allen Systemteilnehmern, doch der entstehende Verwaltungsaufwand und die Notwendigkeit, wiederum eine zentrale Instanz einzurichten, machen diese Variante unattraktiv.
- Schließlich hat das Recht Stellvertretungsregeln geschaffen, die das Handeln eines Systems – einer Gesellschaft – erst möglich machen. So können die Organe einer Gesellschaft für diese handeln. Dies entspricht jedoch nicht der Situation in einem Peer-to-Peer-System, in denen lediglich gleichberechtigte Teilnehmer vorhanden sind.

7.4.1.4 Klimaschutz

Bemerkenswert sind auch die Bemühungen um eine Lösung der Problematik des Zusammenwirkens von Systemteilnehmern im Bereich des Klimaschutzes. Hier zeigt sich eine Form des Gefangenendilemmas: Individuell ist es für jeden Teilnehmer des Systems rational, nichts in die Reduktion von klimaschädlichen Emissionen zu investieren, da der Beitrag einer einzelnen Person oder eines einzelnen Unternehmens – und im wesentlichen auch der Beitrag eines einzelnen Landes – zu klein ist, um spürbare Auswirkungen auf das Weltklima zu haben. Das Systemverhalten insgesamt wird allerdings durch die Summe aller Emissionen bestimmt; bei einer globalen Betrachtung ist wiederum die Reduktion von Emissionen rational.

Auch, wenn im Bereich des Klimaschutzes nun noch keine endgültige Lösung gefunden wurde, so erscheint doch der im Rahmen des Kyoto-Protokolls gewählte marktbasierte Ansatz interessant. Idee des Verfahrens ist, die Kosten, die für das Gesamtsystem entstehen, durch künstliche Schaffung eines Marktes auch in die Kalkulation einzelner Teilnehmer einfließen zu lassen. In der Terminologie der Wirtschaftswissenschaften spricht man hier von einer Internalisierung externer Kosten. Konkret wird ein Markt für das Recht, eine bestimmte Menge CO_2 auszustoßen, geschaffen.

Denkbar wäre ein marktbasierter Ansatz auch für Peer-to-Peer-Systeme. Im Fall des Klimaschutzes wurde dabei der Ansatz gewählt, eine Ressource künstlich zu verknappen und somit zu erreichen, dass diese einen (positiven) Preis erhält. Mit dem Marktsystem wäre somit ein dezentral funktionierender Mechanismus gefunden, der die Auswirkungen des Gesamtsystems auf die Aktionen einzelner Teilnehmer zurückführen kann. Problematisch an diesem Ansatz ist jedoch

- das Finden einer Größe, die stellvertretend für den Beitrag eines einzelnen Knotens zum System herangezogen werden kann – beispielsweise die übertragene Datenmenge.
- das Schaffen einer Infrastruktur, die Zahlungsströme (ggf. auch kleiner Geldbeträge) im System ermöglicht.
- die Notwendigkeit einer Einrichtung, die eventuell geltend gemachte Ansprüche abwehren oder befriedigen kann.

Der Aufwand, der zur Überwindung dieser Probleme nötig wäre, lässt die Verwirklichung eines solchen Ansatzes zumindest in näherer Zukunft nicht als machbar erscheinen. Doch wie kann das Rechtssystem sonst auf Peer-to-Peer-Systeme steuernd einwirken? Diese Frage wird in den folgenden Abschnitten diskutiert.

7.4.2 Einwirkungsmöglichkeiten des Rechts

Wie oben gezeigt, liegt ein wesentliches Problem der Regulierung von Peer-to-Peer-Systemen in der Diffusion von Verantwortlichkeit. Der Gesetzgeber könnte dieser prinzipiell auf verschiedene Arten begegnen:

Sanktionierung rechtswidrigen Handelns Auch, wenn ein einzelner Nutzer eines Peer-to-Peer-Systems rechtswidrig handelt, ist die Wahrscheinlichkeit gering, dass diese Handlung sanktioniert werden kann. Um dennoch einen Anreiz zu setzen, die Handlung zu unterlassen, wäre es also nötig, entsprechend hohe Sanktionen zu wählen. Für den Einzelnen, dessen Verhalten sanktioniert wird, wären solche Sanktionen – seien sie straf- oder zivilrechtlicher Natur – indes nicht angemessen.

Schadensersatzpflichten Ein ähnliches Problem stellt sich bezüglich der Verpflichtung, Schadensersatz zu leisten. Selbst wenn ein Schaden einem einzelnen Nutzer als Verursacher zugerechnet werden kann, ist es denkbar, dass die Verfolgung eines Schadensersatzanspruchs zu hohe Transaktionskosten erfordert, so dass sie unterbleibt. Diesen Fällen durch Erhöhen der Schadensansprüche über den tatsächlich entstandenen Schaden hinaus – ähnlich den „Punitive Damages" [Schw03b] in den USA – zu begegnen, wäre zwar denkbar, käme aber wiederum einer unangemessenen Sanktionierung des einzelnen Schädigers gleich.

Durchsetzung von Unterlassungsansprüchen Als Beispiel für diese Problematik kann eine Objektbewertung gelten, die unwahre Tatsachenbehauptungen in Textform beinhaltet. Ist der Urheber nicht feststellbar, kann der Betroffene dennoch von den speichernden Knoten Unterlassung verlangen. Dies ist bereits nach derzeitiger Rechtslage der Fall (siehe dazu Abschnitt 4.3.5.3 sowie [Stad05, Rn. 26] und [BGH04]). Allerdings kann dieser Lösungsweg technisch erheblich erschwert werden; das ist schon dann der Fall, wenn Inhalte auf mehreren Knoten repliziert gespeichert werden.

Rückführung auf verfolgbare Entitäten Eine mögliche Lösung liegt auch darin, die Verantwortung für ein Peer-to-Peer-System auf Entitäten zurückzuführen, die einer Verfolgung prinzipiell zugänglich sind. Konkret kämen dazu die Ersteller der Software in Frage, die das Peer-to-Peer-System implementiert. Auch diese Lösung ist jedoch problematisch:

- Typischerweise erlauben Peer-to-Peer-Protokolle sowohl rechtmäßige als auch rechtswidrige Nutzungen, wie dies im Fall des Filesharing bisher am deutlichsten offensichtlich wurde. Der Ersteller einer Software kann den konkreten Einsatz oft nicht absehen, eine Verantwortlichkeit dafür lässt sich also nur schwer begründen. Denkbar ist allenfalls eine Lösung nach dem Vorbild des § 95a Abs. 3 des Urheberrechtsgesetzes (UrhG). Dieser lautet:

  > Verboten sind die Herstellung, die Einfuhr, die Verbreitung, der Verkauf, die Vermietung, die Werbung im Hinblick auf Verkauf oder Vermietung und der gewerblichen Zwecken dienende Besitz von Vorrichtungen, Erzeugnissen oder Bestandteilen sowie die Erbringung von Dienstleistungen, die
  >
  > 1. Gegenstand einer Verkaufsförderung, Werbung oder Vermarktung mit dem Ziel der Umgehung wirksamer technischer Maßnahmen sind oder
  > 2. abgesehen von der Umgehung wirksamer technischer Maßnahmen nur einen begrenzten wirtschaftlichen Zweck oder Nutzen haben oder
  > 3. hauptsächlich entworfen, hergestellt, angepasst oder erbracht werden, um die Umgehung wirksamer technischer Maßnahmen zu ermöglichen oder zu erleichtern.

Eine ähnliche Vorschrift ließe sich auch für Peer-to-Peer-Systeme finden – so könnte die Verbreitung von Peer-to-Peer-Software verboten werden, wenn diese unter Hinweis auf die Ermöglichung von Rechtsverstößen beworben wird. Dies gilt zumindest, soweit diese keine Identifikation für einzelne Rechtsverstöße konkret verantwortlicher Personen zulassen. Ein solches Vorgehen wäre jedoch problematisch. Bereits die Regelung des § 95a UrhG rief verfassungsrechtliche Bedenken hervor. Betroffen ist aus Sicht der Nutzer insbesondere[10] das Grundrecht auf Informationsfreiheit aus Art. 5 Abs. 1 Satz 1 (2. Halbsatz) GG. Demnach hat jeder das Recht, „sich aus allgemein zugänglichen Quellen ungehindert zu unterrichten". Auch die Speicherung und Vervielfältigung digitaler Medien ist durch dieses Grundrecht geschützt, die Verbote der §§ 95a Abs. 1,3 demnach ein Eingriff in die Informationsfreiheit der Nutzer [HoBr03, S. 769, mit weiteren Nachweisen]. Zwar wird § 95a UrhG in der Literatur dennoch überwiegend als verfassungskonform angesehen[11], doch würde sich die Lage bei einem allgemeinen Verbot anonymitätsunterstützender und zensurresistenter Peer-to-Peer-Systeme anders darstellen:

– Peer-to-Peer-Systeme sind in vielen Fällen zur Verbreitung von Informationen bestimmt. Anders als im Fall des Urheberrechts, in dem der Urheber den Schutz des § 95a UrhG in Anspruch nehmen kann, aber nicht muss, wären von einer entsprechenden Vorschrift nicht nur Konsumenten der Informationen betroffen, sondern auch diejenigen, die diese zu verbreiten wünschen. Es kommt also zusätzlich ein Verstoß gegen das Recht auf freie Meinungsäußerung (Art. 5 Abs. 1 Satz 1 (1. Halbsatz) GG) in Frage. Dies gilt zumindest, wenn durch das Peer-to-Peer-System eine Meinung verbreitet werden kann, wie dies beispielsweise in dem Empfehlungssystem der Fall ist, das in dieser Arbeit entworfen wurde. Fraglich ist jedoch zunächst, ob Peer-to-Peer-Systeme überhaupt von der Formulierung des Art. 5 Abs. 1 Satz 1 (1. Halbsatz) GG umfasst sind, werden doch nur „Wort, Schrift und Bild" für die Verbreitung der Meinung aufgezählt. Nun lässt sich argumentieren, mit „Wort" sei nur das gesprochene Wort gemeint[12], da ansonsten die Erwähnung der „Schrift" in der Aufzählung nicht nötig wäre. Auch das Vorliegen einer Schrift selbst könnte verneint werden; so erfüllt elektronisch übermittelter Text im bürgerlichen Recht beispielsweise in der Regel nicht die Schriftform. Mit Starck [Star05, Rn. 30 zu Artikel 5 GG] kann jedoch von einem im Verfassungsrecht weitergehenden Schriftbegriff ausgegangen werden, der auch

[10] Neben dem Eigentumsrecht aus Art. 14 Abs. 1 Satz 1 GG, das vorliegend jedoch nicht betrachtet werden soll.

[11] Anderer Auffasung jedoch Ulbricht [Ulbr04, S. 678 f.], der davon ausgeht, dass der Grundrechtseingriff nicht erforderlich und somit nicht verhältnismäßig ist.

[12] Die Auslegung des Begriffs als gesprochenes, ggf. auch aufgezeichnetes Wort wird so auch vertreten von Starck in [Star05, Rn. 29 zu Artikel 5 GG], Herzog in [HSHK07, Rn. 70 zu Artikel 5 GG, Lfg. 30 vom Dezember 1992]

in elektronischer Form übermittelte Schriften abdeckt. Ohnehin ist die Aufzählung von „Wort, Schrift und Bild“ nicht erschöpfend[13].

– Zugleich könnte auch ein Eingriff in das Recht auf informationelle Selbstbestimmung aus Art. 1 Abs. 1 i.V.m. Art. 2 Abs. 1 GG vorliegen: Teilnehmer könnten sich dazu gezwungen sehen, Informationen unter Inkaufnahme eines schlechteren Schutzes ihrer Identität abzurufen oder zu veröffentlichen. Auf diesem Weg würden vermeidbare personenbezogene Daten entstehen.

In der Summe stehen einem Verbot anonymitätsunterstützender und zensurresistenter Peer-to-Peer-Systeme also erhebliche verfassungsrechtliche Bedenken entgegen. Ob diese tatsächlich zur Verfassungswidrigkeit eines solchen Verbots führen würden, ist eine Frage der Abwägung. Wird eine Software, die den Aufbau eines Peer-to-Peer-Systems ermöglicht, unter Hinweis auf die Ermöglichung von Rechtsverstößen beworben und hat sie außer diesem rechtswidrigen Einsatzzweck nur einen begrenzten anderen Nutzen, so dürfte ein Verbot unproblematisch sein. Davon kann als Regelfall jedoch nicht ausgegangen werden.

Es zeigt sich somit, dass die Handlungsmöglichkeiten des Gesetzgebers bei der Regulierung von Peer-to-Peer-Systemen insgesamt sehr beschränkt sind.

7.5 Fazit

In den letzten Abschnitten wurden Lösungsmöglichkeiten für einen Interessenausgleich der Nutzer von Peer-to-Peer-Systemen einerseits und der durch diese Systeme in ihren Rechten verletzten Personen andererseits gesucht. Es stellt sich dabei jedoch heraus, dass eine befriedigende Lösung des Problems durch die gefundenen Lösungsansätze nicht möglich ist. Solange eingesetzte Peer-to-Peer-Systeme die Identität ihrer Nutzer nicht schützen, ist zwar eine Verfolgung einzelner Nutzer möglich, sofern diese rechtswidrig handeln. Ein Peer-to-Peer-System, das die Identität seiner Nutzer auf technischem Wege schützt, lässt dies jedoch nicht mehr zu. Praktikabel lassen sich nur Auswüchse – wie das Bewerben von Software mit dem Argument möglicher Rechtsverletzungen – bekämpfen.

Das heißt jedoch nicht, dass das Rechtssystem Rechtsverletzungen in Peer-to-Peer-Systemen gegenüber machtlos ist. Zwar können anonymitätsunterstützende Peer-to-Peer-Systeme schon mit Blick auf das Informationelle Selbstbestimmungsrecht nicht verboten werden. Demnach ist es auch nicht möglich, Rechtsverletzungen in allen Fällen zu sanktionieren oder Unterlassungs- und Schadensersatzansprüche geltend zu machen. Diese Situation ist allerdings keine Besonderheit solcher Peer-to-Peer-Systeme. Sie treten auch bei ein-

[13] Diese h.M. wird mit einer historischen Auslegung der Norm begründet; siehe nur Herzog in [HSHK07, Rn. 73 zu Artikel 5 GG, Lfg. 30 vom Dezember 1992], Starck in [Star05, Rn. 28], Hoffmann-Riem in [Bäum89, Rn. 25 zu Artikel 5 Abs. 1,2 GG] jeweils mit weiteren Nachweisen.

fachen Peer-to-Peer-Systemen, die die Identität der Nutzer nicht schützen und bei Diensten, die auf Client/Server-Basis angeboten werden, auf. Allein die grenzüberschreitende Kommunikation, die über das Internet ermöglicht wird, macht die Verfolgung von Rechtsverletzungen oft unmöglich. Selbst außerhalb des Internets treten vergleichbare Probleme auf, denn auch in diesem Fall kann nicht davon ausgegangen werden, dass bei jeder Rechtsverletzung der Verursacher ermittelt werden kann.

Nun lässt sich einwenden, durch Peer-to-Peer-Systeme, die die Identität ihrer Nutzer verstecken, werde eine neue Qualität dieser Problematik erreicht, weil ein Nutzer, der andere schädigen will, in diesen Systemen sicher sein könne, nicht zur Rechenschaft gezogen zu werden. Diese Einschätzung kann indes nicht geteilt werden. Dies lässt sich am Beispiel der Empfehlungssysteme begründen. Zu deren Unterstützung der Einsatz von Reputationssystemen geboten, die missbräuchliche Bewertungen bereits deutlich weniger attraktiv machen können. Reputationssysteme erfordern außerdem eine Identifizierung der Nutzer, also zumindest Pseudonyme. Wer nun trotzdem rechtswidrige Bewertungen verfasst, setzt sich dem Risiko aus, verfolgt werden zu können – zwar nicht durch eine Identifizierung auf technischem Wege, aber durch die Inhalte verfasster Bewertungen.

Es bleibt also festzuhalten, dass der Einsatz von Peer-to-Peer-Systemen, die die Identität ihrer Nutzer besser schützen als bislang, zwar eine Erschwernis für die Verfolgung von Rechtsverletzungen ist, sich daraus aber nicht folgern lässt, die Rolle des Rechts müsse sich angesichts neuer technischer Entwicklungen grundlegend ändern.

Kapitel 8

Zusammenfassung und Ausblick

Die vorliegende Arbeit betrachtet selbstorganisierende Empfehlungssysteme, wie sie im Internet eingesetzt werden können, aus Sichtweise der Informatik und der Rechtswissenschaft.

Als wesentliche Herausforderung solcher Systeme kann der Datenschutz gelten: Empfehlungssysteme sollen Nutzern personalisierte Empfehlungen für Objekte bzw. Produkte geben, die für sie von Interesse sein könnten. Dieser Personalisierung erfordert jedoch die Verarbeitung personenbezogener Daten. Dies gilt insbesondere, falls – wie im Rahmen der vorliegenden Arbeit – der Ansatz des kollaborativen Filterns verfolgt wird. Hier bilden explizite oder implizite Bewertungen durch andere Nutzer die Grundlage der Empfehlungsberechnung. Die Tatsache, dass eine Person ein Objekt benutzt hat, stellt aber bereits eine personenbezogene Information dar.

Datenschutzanforderungen, wie sie sich aus Sicht der Nutzer ergeben können, lassen sich in ein Modell aus fünf Ebenen (Zweck-, Kontroll-, Inhalts-, Identitäts- und Empfängerebene) einordnen. Da die vorhandene juristische Literatur sich im Zusammenhang mit Peer-to-Peer-Netzen fast ausschließlich mit der Anwendung des „Filesharing“ – und in der Folge mit urheberrechtlichen Problemen – befasst, ergab sich die Notwendigkeit einer grundlegende Aufarbeitung der juristischen Einordnung solcher Netze.

Die Betrachtung führt zu dem Ergebnis, dass ein Peer-to-Peer-Netz insgesamt als Teledienst im Sinne von § 2 Abs. 1 Teledienstegesetz (TDG) bzw. nach neuer Rechtslage als Telemedium im Sinne von § 1 Abs. 1 Telemediengesetz (TMG) eingeordnet werden kann, was jedoch ohne praktische Konsequenzen bleibt. Relevant ist aber die Einordnung der einzelnen Teilnehmer. Trotz des geringen Umfangs der durch einen einzelnen Teilnehmer jeweils erbrachten Dienstleistung ist dieser als Diensteanbieter im Sinne von TDG und Teledienstedatenschutzgesetz (TDDSG) bzw. im Sinne des TMG zu qualifizieren. Ihn treffen somit (neben den sich aus dem Bundesdatenschutzgesetz ohnehin ergebenden Pflichten) auch die datenschutzrechtlichen Pflichten aus TDDSG und TMG. Während die Verwendung personenbezogener Daten eines Nutzers durch das Gesetz erlaubt wird, sofern dies nötig ist, um den Dienst für diesen Nutzer zu erbringen, erfordert die Verwendung, um einen Dienst für andere Nutzer zu erbringen, eine Einwilligung. Alternativ kann der Personenbezug durch

Anonymisierung oder eine geeignete Pseudonymisierung entfernt werden. Wie die Analyse gezeigt hat, sind die durch das Recht gestellten Datenschutzanforderungen aus Sicht der Informatik durchaus zu erfüllen; entsprechende Vorkehrungen sind aber in bisher implementierten Peer-to-Peer-Systemen meist nicht vorgesehen.

Eng mit dem Datenschutz verknüpft ist die Fragestellung des Vertrauens. So kann die Entscheidung, eigene personenbezogene Daten preiszugeben, von dem Vertrauen abhängen, das eine Person in den Empfänger dieser Daten hat. Vertrauen und insbesondere Systeme, die dazu dienen können, Vertrauen in andere Nutzer herzustellen (Reputationssysteme) wurden daher zunächst juristisch betrachtet. Grundsätzlich sind die Bewertungen – auch solche, die Vertrauen in andere Personen ausdrücken – durch das Recht auf freie Meinungsäußerung (Artikel 5 Abs. 1 GG) geschützt. Werden die Grenzen der freien Meinungsäußerung überschritten, so können sich Unterlassungs- und Schadensersatzansprüche insbesondere aus §§ 1004, 823 Abs.1,2, 824 BGB sowie aus wettbewerbsrechtlichen Normen ergeben. Die Ergebnisse lassen sich auch auf Empfehlungssysteme übertragen.

Die Verknüpfung von Vertrauen und Datenschutz wurde sodann operationalisiert und ein Verfahren entworfen, mit dem Vertrauensbeziehungen durch Verwendung kryptographischer Verfahren auch für Zwecke des Datenschutzes verwendet werden können. Vorausgesetzt wird dabei das Vorhandensein einer Public-Key-Infrastruktur und gewisser Vertrauensbeziehungen zwischen Nutzern. Grundidee des Verfahrens ist, die Vertraulichkeit eines Dokuments als das Maß an Vertrauen auszudrücken, das der Autor einem potentiellen Empfänger entgegenbringen muss, soll diesem die Möglichkeit gegeben werden, das Dokument zu lesen. Die Vertraulichkeitseinstufung wird jedem Dokument beigefügt; ein protokollkonformer Netzknoten darf das Dokument nur weiterleiten, wenn er anhand eines gegebenen Vertrauensmodells einen Vertrauenspfad zwischen Autor des Dokuments und potentiellem Empfänger herstellen kann. Die Herstellung dieses Pfades wird erleichtert, indem beim Versand eines Dokuments eine Liste vertrauter Knoten hinzugefügt wird. Jeder Eintrag wird mit dem öffentlichen Schlüssel des jeweiligen Empfängers verschlüsselt, so dass der jeweils vertraute Knoten an einer Offenlegung von Vertrauensbeziehungen mitwirken muss – auch hierbei handelt es sich um personenbezogene Daten.

Im Anschluss wurden Verfahren zur verteilten Empfehlungserzeugung in Peer-to-Peer-Systemen betrachtet. Alle betrachteten Ansätze beruhen auf einem Overlay-Netz, auf dem eine Datenspeicherungsschicht aufgebaut wird. Diese Datenspeicherungsschicht ist die Grundlage dreier Anwendungen: Die verteilte Speicherung von Bewertungsdokumenten, objektbasiertes und nutzerbasiertes kollaboratives Filtern.

Objektbasiertes Kollaboratives Filtern führt dabei zur besten Empfehlungsqualität; jedoch wächst der Kommunikationsaufwand im schlechtesten Fall sowohl beim Einfügen von Bewertungen als auch beim Abruf von Empfehlungen quadratisch mit der Größe der Nutzerprofile – auch, wenn durch die vorgenommenen Optimierungen der Aufwand in prak-

tischen Szenarien deutlich reduziert werden konnte. Im Fall des nutzerbasierten kollaborativen Filterns ist dieser Aufwand konstant, doch kann nur für eine geringe Anzahl an Objekten eine Vorhersage der Bewertung durch den Nutzer erstellt werden. Um die Vorteile beider Ansätze zu vereinen, wurde ein kombiniertes Verfahren entwickelt. Es zeigt sich, dass hierbei der Aufwand für das Einfügen von Empfehlungen im Vergleich zum nutzerbasierten Ansatz nur geringfügig steigt, während die Anzahl erzeugbarer Empfehlungen drastisch zunimmt.

Den Abschluss der Arbeit bilden Betrachtungen zu möglichen Anpassungen des Rechts an die Entwicklung von Peer-to-Peer-Systemen. Konkret werden gesetzgeberische Fehlleistungen aufgezeigt, und es wird dazu angeregt, die Differenzierung zwischen Diensteanbietern und Nutzern im Rahmen des Telemedienrechts zu überdenken.

Literaturverzeichnis

[Abdu97] Alfarez Abdul-Rahman. The PGP Trust Model. *EDI Forum: The Journal of Electronic Commerce* 10(3), April 1997, S. 27–31.

[AcCR99] Mark S. Ackerman, Lorrie Faith Cranor und Joseph Reagle. Privacy in e-commerce: examining user scenarios and privacy preferences. In *EC '99: Proceedings of the 1st ACM conference on Electronic commerce, Denver/Colorado, November 1999*. ACM Press, New York, 1999, S. 1–8.

[AdTu05] Gediminas Adomavicius und Alexander Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering* 17(6), Juni 2005, S. 734–749.

[AGEr04] Zustimmung zur Löschung einer „eBay"-Bewertung. *Zeitschrift für das gesamte Schuldrecht* (9), 2004, S. 359–361. AG Erlangen, Urteil vom 26.05.2004 – Az. 1 C 457/04.

[AGPe05] Pflicht zur Rücknahme einer Bewertung nach eBay-Kauf. *NJW-Rechtsprechungsreport Zivilrecht* 20(4), 2005, S. 275–276. AG Peine, Urteil vom 15.9.2004 – Az. 18 C 234/04.

[Aker70] George A. Akerlof. The Market for „Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics* 84(3), August 1970, S. 488–500.

[ARHa97] Alfarez Abdul-Rahman und Stephen Hailes. A distributed trust model. In *NSPW '97: Proceedings of the 1997 workshop on New security paradigms, Langdale, Cumbria, UK, September 1997*. ACM Press, New York, 1997, S. 48–60.

[AvRZ99] Christopher Avery, Paul Resnick und Richard Zeckhauser. The Market for Evaluations. *The American Economic Review* 89(3), Juni 1999, S. 564–584.

[Bako97] J. Yannis Bakos. Reducing Buyer Search Costs: Implications for Electronic Marketplaces. *Management Science* 43(12), Dezember 1997, S. 1676–1692.

[BaSc04] Salman A. Baset und Henning Schulzrinne. An Analysis of the Skype Peer-to-Peer Internel Telephony Protocol. Technischer Bericht CUCS-039-04, Columbia University, New York, Dezember 2004.

[BeCr92] Nicholas J. Belkin und W. Bruce Croft. Information filtering and information retrieval: two sides of the same coin? *Communications of the ACM* 35(12), 1992, S. 29–38.

[Bedn07] Mark Bedner. Haftung des Betreibers von Internetforen. *JurPC: Internet-Zeitschrift für Rechtsinformatik und Informationsrecht* (94), 2007. JurPC Web-Dokument 94/2007, `http://www.jurpc.de/aufsatz/20070094.htm`, abgerufen am 15.7.2007.

[BeKi99] Uwe-Christian Behrens und Matthias Kirspel. *Grundlagen der Volkswirtschaftslehre: Einführung*. Erschienen in der Reihe Managementwissen für Studium und Praxis. Oldenbourg, München, Wien. 1999.

[BEKR06] Shlomo Berkovsky, Yaniv Eytani, Tsvi Kuflik und Francesco Ricci. Hierarchical Neighborhood Topology for Privacy Enhanced Collaborative Filtering. In Alfred Kobsa, Ramnath K. Chellappa und Sarah Spiekermann (Hrsg.), *Proceedings of the CHI 2006 Workshop on Privacy-Enhanced Personalization, PEP2006, Montréal, Quebec, 22 April 2006*, 2006, S. 6–13.

[BGH04] BGH: Internetversteigerung – ROLEX. *Multimedia und Recht* 7(10), 2004, S. 668–673. BGH, Urteil vom 11.3.2004 – Az. I ZR 304/01 (OLG Köln, LG Köln).

[BGH88] Haftung der öffentlichen Hand für Waldschäden. *Neue Juristische Wochenschrift* 41(8), 1988, S. 478–482. BGH, Urteil vom 10-12-1987 – III ZR 220/86 (Stuttgart).

[BGH89] Keine Auskunftspflicht zwischen Miterben über Testierfähigkeit des Erblassers. *NJW-Rechtsprechungsreport Zivilrecht* 5(8), 1989, S. 450–451. BGH, Urteil vom 07-12-1988 – Az. IV a ZR 290/87 (Schleswig).

[Bize07] Johann Bizer. Was sind Telemedien? *Datenschutz und Datensicherheit* 31(1), 2007, S. 40.

[Büll99] Alfred Büllesbach. Das TDDSG aus Sicht der Wirtschaft. *Datenschutz und Datensicherheit* 23(5), 1999, S. 263–265.

[BMBC+05] Roland Bless, Stefan Mink, Erik-Oliver Blaß, Michael Conrad, Hans-Joachim Hof, Kendy Kutzner und Marcus Schöller. *Sichere Netzwerkkommunikation*. Springer, Berlin, Heidelberg. Juni 2005.

[Brai98] Valerie Braithwaite. *Trust and Governance*, Band 1 der Reihe *The Russel Sage Foundation Series on Trust*, Kapitel 3 (Communal and Exchange Trust Norms: Their Value Base and Relevance to Institutional Trust), S. 46–74. Russel Sage Foundation, New York. 1998.

[Brin06] Guido Brinkel. *Filesharing*. Nr. 4 der Reihe Geistiges Eigentum und Wettbewerbsrecht. Mohr Siebeck, Tübingen. 2006.

[Brox05] Hans Brox. *Allgemeiner Teil des BGB*. Erschienen in der Reihe Academia Iuris: Lehrbücher der Rechtswissenschaft. Carl Heymanns Verlag, Köln. 29. Auflage, 2005.

[Brun04] Phillip W. Brunst. Umsetzungsprobleme der Impressumspflicht bei Webangeboten. *Multimedia und Recht* 7(1), 2004, S. 8–13.

[Bäum89] Richard Bäumlin et al. *Kommentar zum Grundgesetz für die Bundesrepublik Deutschland*. Erschienen in der Reihe Alternativkommentare. Luchterhand. 2. Auflage, 1989.

[Bund83] Bundesverfassungsgericht. Volkszählung. *Entscheidungen des Bundesverfassungsgerichts* 65(1), 1983.

[Bund90] Deutscher Bundestag. Gesetzentwurf der Bundesregierung: Entwurf eines Umwelthaftungsgesetzes – UmweltHG, 1990. Deutscher Bundestag, Drucksache 11/7104.

[Bund97] Deutscher Bundestag. Entwurf eines Gesetzes zur Regelung der Rahmenbedingungen für Informations- und Kommunikationsdienste (Informations- und Kommunikationsdienstegesetz – IuKDG), 1997. Deutscher Bundestag, Drucksache 13/7385.

[Bund99] Deutscher Bundestag. Bericht der Bundesregierung über die Erfahrungen und Entwicklungen bei den neuen Informations- und Kommunikationsdiensten im Zusammenhang mit der Umsetzung des Informations- und Kommunikationsdienste-Gesetzes (IuKDG), 1999. Deutscher Bundestag, Drucksache 14/1191.

[Bund01a] Deutscher Bundestag. Beschlussempfehlung und Bericht des Ausschusses für Wirtschaft und Technologie (9. Ausschuss) zu dem Gesetzentwurf der Bundesregierung – Drucksache 14/6098 – Entwurf eines Gesetzes über rechtliche Rahmenbedingungen für den elektronischen Geschäftsverkehr (ElektronischerGeschäftsverkehr-Gesetz – EGG), 2001. Deutscher Bundestag, Drucksache 14/7345.

[Bund01b] Deutscher Bundestag. Entwurf eines Gesetzes über rechtliche Rahmenbedingungen für den elektronischen Geschäftsverkehr, 2001. Deutscher Bundestag, Drucksache 14/6098.

[Bund02] Deutscher Bundestag. Stenographischer Bericht: 248. Sitzung, Berlin, Donnerstag, den 4. Juli 2002, 2002. Deutscher Bundestag, Plenarprotokoll 14/248.

[Bund06a] Deutscher Bundestag. Entwurf eines Gesetzes zur Vereinheitlichung von Vorschriften über bestimmte elektronische Informations- und Kommunikationsdienste – (Elektronischer-Geschäftsverkehr-Vereinheitlichungsgesetz – ElGVG), 2006. Deutscher Bundestag, Drucksache 16/3078.

[Bund06b] Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. Daten zur Informationsgesellschaft: Status quo und Perspektiven Deutschlands im internationalen Vergleich. `http://bitkom.de/files/documents/daten_broschuere_2006.pdf`, 2006. Edition 2006, abgerufen am 30.4.2007.

[BVerfG01] Entscheidung des Bundesverfassungsgerichts vom 1.8.2001, 2001. Az. 1 BvR 1906/97.

[BVerfG80] Verhältnis des allgemeinen Persönlichkeitsrechts zur Meinungsfreiheit. *Neue Juristische Wochenschrift* 33(38), 1980, S. 2072–2073. BVerfG, Beschluss vom 3. 6. 1980 – 1 BvR 797/78.

[BVerfG82] Bezeichnung der CSU als "NPD Europas" im Wahlkampf. *Neue Juristische Wochenschrift* 36(25), 1983, S. 1415–1417. BVerfG, Beschluß vom 22.06.1982 – 1 BvR 1376/79.

[Cana71] Claus-Wilhelm Canaris. *Die Vertrauenshaftung im deutschen Privatrecht*, Band 16 der Reihe *Münchener Universitätsschriften: Reihe der Juristischen Fakultät*. C.H. Beck'sche Verlagsbuchhandlung, München. 1971.

[CDGR+02] Miguel Castro, Peter Druschel, Ayalvadi Ganesh, Antony Rowstron und Dan S. Wallach. Secure routing for structured peer-to-peer overlay networks. In *OSDI '02: Proceedings of the 5th symposium on Operating systems design and implementation, Boston, Massachussets, December 2002*. ACM Press, New York, 2002, S. 299–314.

[CDHS+05] Michael Conrad, Jochen Dinger, Hannes Hartenstein, Marcus Schöller und Martina Zitterbart. Combining Service-Orientation and Peer-to-Peer Networks. In Paul Müller, Reinhard Gotzhein und Jens B. Schmitt (Hrsg.), *Kommunikation in Verteilten Systemen (KiVS), 14. ITG/GI-Fachtagung Kommunikati-*

on in Verteilten Systemen (KiVS 2005) Kaiserslautern, 28. Februar - 3. März 2005. Springer, 2005, S. 181–184.

[CDVP+02] Fabrizio Cornelli, Ernesto Damiani, Sabrina De Capitani di Vimercati, Stefano Paraboschi und Pierangela Samarati. Choosing reputable servents in a P2P network. In *WWW '02: Proceedings of the 11th international conference on World Wide Web*, New York, NY, USA, 2002. ACM Press, New York, S. 376–386.

[Cede07] Jörgen Cederlöf. Wotsap: Web of trust statistics and pathfinder. `http://webware.lysator.liu.se/jc/wotsap/wots/latest/wotinfo.txt`, abgerufen am 13.07.2007, 2007.

[CGBD+05] Geoff Coulson, Paul Grace, Gordon Blair, David Duce, Chris Cooper und Musbah Sagar. A Middleware Approach for Pervasive Grid Environments. In *UK-UbiNet/ UK e-Science Programme Workshop on Ubiquitous Computing and e-Research, Edinburgh, UK, 18-19 May 2005*, 2005. Elektronische Veröffentlichung.

[Chau81] David Chaum. Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. *Communications of the ACM* 24(2), Februar 1981, S. 84–88.

[ClKe66] Cyril W. Cleverdon und Michael Keen. *Aslib Cranfield Research Project: Factors determining the Performance of Indexing Systems*, Band 2. Cyril Cleverdon, Cranfield. 1966.

[CoFa00] William W. Cohen und Wei Fan. Web-collaborative filtering: recommending music by crawling the Web. *Computer Networks* 33(1–6), 2000, S. 685–698.

[Cole90] James Samuel Coleman. *Foundations of Social Theory*. The Bellknap Press of Harvard University Press, Cambridge/London. 1990.

[CrRA99] Lorrie Faith Cranor, Joseph Reagle und Mark S. Ackerman. Beyond concern: Understanding net users' attitudes about online privacy. Research Technical Report TR 99.4.3, AT&T Labs, 1999.

[CSWH01] Ian Clarke, Oskar Sandberg, Brandon Wiley und Theodore W. Hong. Freenet: A Distributed Anonymous Information Storage and Retrieval System. In H. Federrath (Hrsg.), *Designing Privacy Enhancing Technologies: International Workshop on Design Issues in Anonymity and Unobservability, Berkeley, CA, USA, July 2000, Proceedings*, Band 2009 der Reihe *Lecture Notes in Computer Science*, Berlin, Heidelberg, 2001. Springer, S. 46–66.

[Darm48] Friedrich Darmstaedter. Recht und Jurist. *Süddeutsche Juristenzeitung* Band 3, 1948, Spalten 430-436.

[Dasg88] Partha Dasgupta. *Trust: making and breaking cooperative relations*, Kapitel 4: Trust as a Commodity, S. 49–72. Basil Blackwell, Oxford. 1988.

[DBKK+01] Frank Dabek, Emma Brunskill, M. Frans Kaashoek, David Karger, Robert Morris, Ion Stoica und Hari Balakrishnan. Building Peer-to-Peer Systems with Chord, a Distributed Lookup Service. In *Proceedings of HotOS-VIII: 8th Workshop on Hot Topics in Operating Systems, May 20-23, 2001, Elmau/Oberbayern, Germany*, Washington, DC, USA, 2001. IEEE Computer Society, S. 81–86.

[DeKa04] Mukund Deshpande und George Karypis. Item-Based-Top-N Recommendation Algorithms. *ACM Transactions on Information Systems* 22(1), Januar 2004, S. 143–177.

[DiHa06] Jochen Dinger und Hannes Hartenstein. Defending the Sybil Attack in P2P Networks: Taxonomy, Challenges, and a Proposal for Self-Registration. In *First International Conference on Availability, Reliability and Security (ARES'06), 20th-22nd April 2006, Vienna University of Technology*. IEEE Computer Society, Los Alamitos, CA, USA, 2006, S. 756–763.

[DöKo07] Tanja Dörre und Kai Kochmann. Zivilrechtlicher Schutz gegen negative eBay-Bewertungen. *Zeitschrift für Urheber- und Medienrecht* 51(1), 2007, S. 30–40.

[Douc02] John R. Douceur. The Sybil Attack. In P. Druschel, F. Kaashoek und A. Rowstron (Hrsg.), *Peer-to-Peer Systems: First International Workshop, IPTPS 2002, Cambridge, MA, USA, March 7-8, 2002, Revised Papers*, Band 2429 der Reihe *Lecture Notes in Computer Science*, 2002.

[Drom97] R. Droms. Dynamic Host Configuration Protocol. RFC 2131 (Draft Standard), März 1997. Updated by RFC 3396.

[DuGH03] Schahram Dustdar, Harald Gall und Manfred Hauswirth. *Software-Architekturen für Verteilte Systeme*, Kapitel 7: Peer-to-Peer-Architekturen, S. 161–197. Xpert.press. Springer, Berlin, Heidelberg, New York. 2003.

[EBAY07] Allgemeine Geschäftsbedingungen für die Nutzung der deutschsprachigen eBay-Websites. `http://pages.ebay.de/help/policies/user-agreement.html?ssPageName=f:f:DE`, 2007. Abgerufen am 6. Juli 2007.

[EFMT97] Stefan Engel-Flechsig, Frithjof A. Maennel und Alexander Tettenborn. Das neue Informations- und Kommunikationsdienste-Gesetz. *Neue Juristische Wochenschrift* 50(45), 1997, S. 2981–2992.

[EFMT98] Stefan Engel-Flechsig, Frithjof A. Maennel und Alexander Tettenborn. *Neue gesetzliche Rahmenbedingungen für Multimedia*. Verlag Recht und Wirtschaft, Heidelberg. Sonderveröffentlichung der Zeitschrift *Betriebs-Berater*, 1998.

[EFMT01] Stefan Engel-Flechsig, Frithjof A. Maennel und Alexander Tettenborn (Hrsg.). *Beck'scher IuKDG-Kommentar*. C.H. Beck, München. 2001.

[Enge97] Stefan Engel-Flechsig. Das Informations- und Kommunikationsdienstegesetz des Bundes und der Mediendienstestaatsvertrag der Bundesländer. *Zeitschrift für Urheber- und Medienrecht* 41(4), 1997, S. 231–239.

[Erns04a] Stefan Ernst. *[Erns04b]*, Kapitel Strafrechtliche Fragen, S. 79–146. 2004.

[Erns04b] Stefan Ernst (Hrsg.). *Hacker, Cracker & Computerviren: Recht und Praxis der Informatinssicherheit*. Verlag Dr. Otto Schmidt, Köln. 2004.

[Esse00] Hartmut Esser. *Institutionen*, Band 5 der Reihe *Soziologie: Spezielle Grundlagen*. Campus Verlag, Frankfurt/New York. 2000.

[Füll05] Michael Fülling. Die Telefonnummer als Pflichtangabe der Anbieterkennzeichnung nach § 6 TDG? *Multimedia und Recht* 8(9), 2005, S. V–VI.

[FMLO+01] B. J. Fogg, Jonathan Marshall, Othman Laraki, Alex Osipovich, Chris Varma, Nicholas Fang, Jyoti Paul, Akshay Rangnekar, John Shon, Preeti Swani und Marissa Treinen. What makes Web sites credible? A report on a large quantitative study. In Michel Beaudouin-Lafon und Robert J. K. Jacob (Hrsg.), *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems, Seattle, Washington, USA, March 31 - April 5, 2001*. ACM Press, New York, 2001, S. 61–68.

[Fox00] Susannah Fox. Trust and Privacy Online: Why Americans Want to Rewrite the Rules. Studie, Pew Internet & American Life Project, Washington, D.C., August 2000.

[FrRe01] Eric Friedman und Paul Resnick. The Social Cost of Cheap Pseudonyms. *Journal of Economics and Management Strategy* 10(2), 2001, S. 173–199.

[FSCM02] Michael J. Freedman, Emil Sit, Josh Cates und Robert Morris. Introducing Tarzan, a Peer-to-Peer Anonymizing Network Layer. In Peter Druschel, M. Frans Kaashoek und Antony I. T. Rowstron (Hrsg.), *IPTPS*, Band 2429 der Reihe *Lecture Notes in Computer Science*. Springer, 2002, S. 121–129.

[Gamb88] Diego Gambetta. *Trust: Making and Breaking Cooperative Relations*, Kapitel Can we trust trust?, S. 213–237. Basil Blackwell, Oxford. 1988.

[GLSS00] Edward L. Glaeser, David I. Laibson, José A. Scheinkman und Christine L. Soutter. Measuring Trust. *The Quarterly Journal of Economics* 115(3), 2000, S. 811–846.

[GNOT92] David Goldberg, David Nichols., Brian M. Oki und Douglas Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM* 35(12), 1992, S. 61–70.

[GoKl03] Peter Gola und Christoph Klug. *Grundzüge des Datenschutzrechts*. C.H. Beck, München. 2003.

[Golb06] Jennifer Golbeck. Generating Predictive Movie Recommendations from Trust in Social Networks. In *Trust Management: 4th International Conference, Itrust 2006, Pisa, Italy, May 16-19, 2006, Proceedings*, Band 3986 der Reihe *Lecture Notes in Computer Science*, Berlin, Heidelberg, 2006. S. 93–104.

[GoMü00] Peter Gola und Thomas Müthlein. *TDG, TDDSG: Teledienstegesetz, Teledienstedatenschutzgesetz – Kommentierung für die Praxis*. Datakontext Fachverlag, Frechen. 2000.

[Goun97] Georgios Gounalakis. Der Mediendienste-Staatsvertrag der Länder. *Neue Juristische Wochenschrift* 50(45), 1997, S. 2993–3000.

[GPMP+05] Pedro García, Carles Pairot, Rubén Mondéjar, Jordi Pujol, Helio Tejedo und Robert Rallo. PlanetSim: A New Overlay Network Simulation Framework. In Thomas Gschwind und Cecilia Mascolo (Hrsg.), *Software Engineering and Middleware: 4th International Workshop, SEM 2004, Linz, Austria, September 20-21, 2004. Revised Selected Papers*, Band 3437 der Reihe *Lecture Notes in Computer Science*, Berlin, Heidelberg, New York, 2005. Springer, S. 123–136.

[GPSS06] Martin Geppert, Hermann-Josef Piepenbrock, Raimund Schütz und Fabian Schuster (Hrsg.). *Beck'scher TKG-Kommentar*. C.H. Beck, München. 3. Auflage, 2006.

[GRGP01] Kenneth Y. Goldberg, Theresa Roeder, Dhruv Gupta und Chris Perkins. Eigentaste: A Constant Time Collaborative Filtering Algorithm. *Information Retrieval* 4(2), 2001, S. 133–151.

[Grub93] Thomas R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition* 5(2), 1993, S. 199–200.

[GuJA03] Minaxi Gupta, Paul Judge und Mostafa Ammar. A reputation system for peer-to-peer networks. In *Proceedings of the 13th international workshop on Network*

and operating systems support for digital audio and video, NOSSDAV 2003, June 1-3, 2003, Monterey, California, USA. ACM Press, New York, 2003, S. 144–152.

[Hage86] Günter Hager. Umweltschäden - ein Prüfstein für die Wandlungs- und Leistungsfähigkeit des Deliktsrechts. *Neue Juristische Wochenschrift* 39(32), 1986, S. 1961–1971.

[Hals96] Fred Halsall. *Data Communications, Computer Networks and Open Systems*. Addison-Wesley, Harlow, Reading, Menlo Park. 4. Auflage, 1996.

[HaRe07] Günter Hager und Eckard Rehbinder. *Sonstiges Umweltrecht*, Band 3, Kapitel 2 (Umwelthaftungsgesetz). C.H. Beck, München. Begründet von Robert von Landmann und Gustav Rohmer; Stand: 50. Ergänzungslieferung, 2007.

[Harr01] Harris Study no. 14875. Technischer Bericht, Harris Interactive, New York, NY/Chapel Hill, NC, August 2001. `www.irss.unc.edu/data_archive/pollsearch.html`, abgerufen am 18.7.2006.

[HeMN98] Robin Helmke, Björn Müller und Andreas Neumann. Internet-Telefonie zwischen TKG, IuKDG und Mediendienste-Staatsvertrag. *JurPC: Internet-Zeitschrift für Rechtsinformatik und Informationsrecht* (93), 1998. JurPC Web-Dokument 93/1998, `http://www.jurpc.de/aufsatz/19980093.htm`, abgerufen am 6.7.2007.

[HKTR04] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen und John T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems* 22(1), 2004, S. 5–53.

[HoBr03] Bernd Holznagel und Sandra Brüggemann. Das Digital Right Management nach dem ersten Korb der Urheberrechtsnovelle: Eine verfassungsrechtliche Beurteilung der neuen Kopierschutzregelungen. *Multimedia und Recht* 6(12), 2003, S. 767–773.

[Hoer98] Thomas Hoeren. Wegerechte auf dem Prüfstand – § 57 TKG und die Nachverlegung von Lichtwellenleiterkabeln. *Multimedia und Recht* 1(1), 1998, S. 1–6.

[HSHK07] Roman Herzog, Rupert Scholz, Matthias Herdegen und Hans Klein (Hrsg.). *Grundgesetz: Kommentar*. C.H. Beck. Loseblattsammlung, begründet von Theodor Maunz und Günter Dürig, 2007.

[Hueb03] Ryan Huebsch. Content-Based Multicast: Comparison of Implementation Options. Technischer Bericht UCB/CSD-03-1229, UC Berkeley, Februar 2003.

[HXYS04] P. Han, B. Xie, F. Yang und R. Shen. A scalable P2P recommender system based on distributed collaborative filtering. *Expert Systems With Applications* 27(2), 2004, S. 203–210.

[Imho00] Ralf Imhof. One-to-One-Marketing im Internet – Das TDDSG als Marketinghindernis. *Computer und Recht* 16(2), 2000, S. 110–116.

[Inte94] International Organization for Standardization (ISO). Information Technology – Open Systems Interconnection – Basic Reference Model: The Basic Model. ISO/IEC 7498-1 | ITU-T Recommendation X.200, Juli 1994.

[Inte01] International Organization for Standardization (ISO). Information and documentation – International Standard Recording Code (ISRC). ISO 3901:2001, 2001.

[Inte05] International Organization for Standardization (ISO). Information and documentation – International standard book number (ISBN). ISO 2108:2005, Mai 2005.

[Inte06] International DOI Foundation. The DOI system, 2006. `http://www.doi.org`, abgerufen am 21.12.2006.

[Jäge06] Marc-Alexander Jäger. Verteilte Empfehlungsberechnung in selbstorganisierenden Systemen. Diplomarbeit, Institut für Telematik, Universität Karlsruhe (TH), Dezember 2006.

[JøIB07] Audun Jøsang, Roslan Ismail und Colin Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems* 43(2), 2007, S. 618–644.

[Jone01] Karen Jones. *International Encyclopedia of the Social and Behavioral Sciences*, Band 23, Kapitel Trust: Philosophical Aspects, S. 15917–15922. Elsevier, Amsterdam, Paris, New York. 2001.

[KaSGM03] Sepandar D. Kamvar, Mario T. Schlosser und Hector Garcia-Molina. The Eigentrust algorithm for reputation management in P2P networks. In *WWW '03: Proceedings of the 12th international conference on WorldWide Web (Budapest, Hungary, May 20 - 24, 2003)*. ACM Press, New York, 2003, S. 640–651.

[Kaza05] Wassilios Kazakos. *Kontextkonforme Empfehlungen auf der Grundlage verteilter Informationen*. Logos, Berlin. 2005.

[KePl05] Peter Kenning und Hilke Plassmann. NeuroEconomics: An overview from an economic perspective. *Brain Research Bulletin* 67(5), 2005, S. 343–354.

[KiPe03] Michael Kinateder und Siani Pearson. A Privacy-Enhanced Peer-to-Peer Reputation System. In K. Bauknecht, A. M. Tjoa und G. Quirchmayr (Hrsg.), *E-Commerce and Web Technologies, 4th International Conference, EC-Web, Prague, Czech Republic, September 2-5, 2003, Proceedings*, Band 2738 der Reihe *Lecture Notes in Computer Science*, Berlin, Heidelberg, 2003. Springer, S. 206–215.

[KMMH+97] Joseph A. Konstan, Bradley N. Miller, David Maltz, Jonathan L. Herlocker, Lee R. Gordon und John Riedl. GroupLens: applying collaborative filtering to Usenet news. *Communications of the ACM* 40(3), 1997, S. 77–87.

[Koch05] Frank A. Koch. *Internet-Recht: Praxishandbuch zu Dienstenutzung, Verträgen, Rechtsschutz und Wettbewerb, Haftung, Arbeitsrecht und Datenschutz im Internet, zu Links, Peer-to-Peer-Netzen und Domain-Recht, mit Musterverträgen*. Oldenbourg, München, Wien. 2. Auflage, 2005.

[KoNe00] Christian Koenig und Andreas Neumann. Die neue Telekommunikations-Datenschutzverordnung. *Kommunikation und Recht* (9), 2000, S. 417–425.

[KrLG02] David Krane, Laura Light und Diana Gravitch. Privacy On and Off the Internet: What consumers want. Study 15229, Harris Interactive, New York, 2002.

[LAKV+01] Richard D. Lawrence, George S. Almasi, Vladimir Kotlyar, Marisa S. Viveros und Sastry S. Duri. Personalization of Supermarket Product Recommendations. *Data Mining and Knowledge Discovery* 5(1), 2001, S. 11–32.

[Levi01] Margaret Levi. *International Encyclopedia of the Social and Behavioral Sciences*, Band 23, Kapitel Trust, Sociology of, S. 15922–15926. Elsevier, Amsterdam, Paris, New York. 2001.

[LeWi86] Theodor Lenckner und Wolfgang Winkelbauer. Computerkriminalität – Möglichkeiten und Grenzen des 2. WiKG (II). *Computer und Recht* 2(10), 1986, S. 654–661.

[Leyd93] Loet Leydesdorff. Is Society a Self-Organizing System? *Journal of Social and Evolutionary Systems* 16(3), 1993, S. 331–349.

[LGD004] LG Düsseldorf: Pflicht zur Löschung unwahrer eBay-Bewertungen. *Computer und Recht* 20(8), 2004, S. 623–624. LG Düsseldorf, Urteil vom 18.2.2004 – Az. 12 O 6/04, rechtskräftig.

[LGDA05] Speicherung von IP-Adressen. *JurPC: Internet-Zeitschrift für Rechtsinformatik und Informationsrecht* (52), 2006. Landgericht Darmstadt, Urteil vom 07.12.2005 – Az. 25 S 118/2005. JurPC Web-Dokument 52/2006, `http://www.jurpc.de/rechtspr/20060052.htm`, abgerufen am 3.8.2006.

[LGKN04] Falsche Tatsachenbehauptungen im eBay-Bewertungssystem. *NJW-Rechtsprechungsreport Zivilrecht* 19(23), 2004, S. 1635–1637. LG Konstanz, Urteil vom 28.7.2004 – Az. 11 S 31/04.

[LGM06] LG München I: Haftung des Auktionsplattformbetreibers für urheberrechtsverletzende Handlungen. *Multimedia und Recht* 9(5), 2006, S. 332–336. LG München I, Urteil vom 11.1.2006 – Az. 21 O 2793/05.

[Libe07] Michael Libertus. Determinanten der Störerhaftung für Inhalte in Onlinearchiven. *Multimedia und Recht* 10(3), 2007, S. 143–149.

[LiSY03] Greg Linden, Brent Smith und Jeremy York. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Computing* 7(1), 2003, S. 76–80.

[MaEr04] Peter Mankowski und Stefan Ernst. *[Erns04b]*, Kapitel Zivilrechtliche Ansprüche, S. 147–284 (S. 147–236: Peter Mankowski; S. 237–284: Stefan Ernst). 2004.

[Mars94] Stephen Paul Marsh. *Formalising Trust as a Computational Concept*. Dissertation, University of Stirling, 1994.

[Meng04] Anja Mengel. Kontrolle der E-mail- und Internetkommunikation am Arbeitsplatz. *Betriebs-Berater* 59(37), 2004, S. 2014–2021.

[MeSC03] Daniela Memmo, Giovanni Sartor und Gioacchino Quadri di Cardano. Trust, Reliance, Good Faith, and the Law. In Paddy Nixon and Sotirios Terzis (Hrsg.), *Trust Management, First International Conference, iTrust 2003, Heraklion, Crete, Greece, May 28-30, 2002, Proceedings*, Band 2692 der Reihe *Lecture Notes in Computer Science*, Berlin, Heidelberg, 2003. Springer, S. 150–164.

[MiFe01] Anthony D. Miyazaki und Ana Fernandez. Consumer perceptions of privacy and security risks for online shopping. *Journal of Consumer Affairs* 35(1), 2001, S. 27–43.

[MiKR04] Bradley N. Miller, Joseph A. Konstan und John Riedl. PocketLens: Toward a personal recommender system. *ACM Transactions on Information Systems* 22(3), 2004, S. 437–476.

[MoDr05] Hans-Werner Moritz und Thomas Dreier (Hrsg.). *Rechts-Handbuch zum E-Commerce*. Verlag Dr. Otto Schmidt, Köln. 2. Auflage, 2005.

[Moy98] John Moy. OSPF Version 2. RFC 2328 (Standard), April 1998.

[NoTa04] Martin Nolte und Christian J. Tams. Grundfälle zu Art. 5 I 1 GG. *Juristische Schulung* 44(2), 2004, S. 111–113.

[O'He99] Mark O'Connor und Jon Herlocker. Clustering Items for Collaborative Filtering. In *Proceedings of the ACM SIGIR Workshop on Recommender Systems, Berkeley, CA, August 1999*, 1999.

[OkMA04] Toshio Oka, Hiroyuki Morikawa und Tomonori Aoyama. Vineyard: A Collaborative Filtering Service Platform in Distributed Environment. In *2004 Symposium on Applications and the Internet Workshops (SAINT 2004 Workshops), 26-30 January 2004, Tokyo, Japan*, Washington, DC, USA, 2004. IEEE Computer Society, S. 575.

[OLG 06] OLG München, 2006. Urteil v. 21.09.2006, Az. 29 U 2119/06.

[OLGBR04] OLG Brandenburg: Haftung eines Onlineauktionshauses. *Multimedia und Recht* 7(5), 2004, S. 330–334. OLG Brandenburg, Urteil vom 16.12.2003 – Az. 6 U 161/02 (LG Potsdam) (nicht rechtskräftig).

[OLGD02] Haftung für Nachrichten im Internet-Portal – News im Internet. *NJW-Rechtsprechungsreport Zivilrecht* 17(13), 2002, S. 910–912. OLG Düsseldorf, Urteil vom 4.10.2001 – Az. 2 U 48/01.

[OLGH04] Pflichtangaben eines Internetanbieters. *NJW-Rechtsprechungsreport Zivilrecht* 20(15), 2004, S. 1045–1047. OLG Hamm, Urteil vom 17.3.2004 – Az. 20 U 222/03, nicht rechtskräftig.

[OLGK02] OLG Köln: „Steffi Graf". *Multimedia und Recht* 5(8), 2002, S. 548–551. OLG Köln, Urteil vom 28.5.2002 – Az. 15 U 221/01 (LG Köln) (rechtskräftig).

[OLGKo07] OLG Koblenz. „Achtung Betrüger unterwegs!" – Zu Abgrenzung von Meinungsäußerung und Schmähkritik und zur Störerhaftung des Internet-Forenbetreibers. *Medien Internet und Recht* (320), 2007. OLG Koblenz, Beschluss vom 12.07.2007 – Az. 5 O 119/06.

[OLGOL06] OLG Oldenburg: Widerruf einer negativen Bewertung bei eBay. *Multimedia und Recht* 9(8), 2006, S. 556–557. OLG Oldenburg, Urteil vom 3.4.2006 – Az. 13 U 71/05 (LG Oldenburg).

[Opas01] Horst W. Opaschowski. *Der gläserne Konsument: Die Zukunft von Datenschutz und Privatsphäre in einer vernetzten Welt.* B.A.T Freizeit-Forschungsinstitut GmbH, Hamburg. 2. aktualisierte und erweiterte. Auflage, 2001.

[Oram01] Andy Oram (Hrsg.). *Peer-to-Peer: Harnessing the Benefits of a Disruptive Technology*. O'Reilly, Beijing, Cambridge, Farnham. 2001.

[O'Sm05] John O'Donovan und Barry Smyth. Trust in recommender systems. In *Proceedings of the 2005 International Conference on Intelligent User Interfaces, January 10-13, 2005, San Diego, California, USA*. ACM Press, New York, 2005, S. 167–174.

[O'Va04] Charles W. O'Donnell und Vinod Vaikuntanathan. Information Leak in the Chord Lookup Protocol. In *Fourth International Conference on Peer-to-Peer Computing (P2P2004), 25-27 August 2004, Zürich, Switzerland, Proceedings*, Washington, DC, USA, 2004. IEEE Computer Society, S. 28–35.

[P3P06] World Wide Web Consortium. The Platform for Privacy Preferences 1.1 (P3P1.1) Specification. `http://www.w3.org/TR/P3P11/`, 2006. abgerufen am 11.7.2007.

[PaJø03] Mary Anne Patton und Audun Jøsang. *Trust in the Network Economy*, Band 2 der Reihe *evolaris-Schriftenreihe*, Kapitel Technologien und Strategien zum Aufbau von Vertrauen im Electronic Commerce, S. 73–88. Springer, Wien, New York. 2003.

[PaJø04] Mary Anne Patton und Audun Jøsang. Technologies for Trust in Electronic Commerce. *Electronic Commerce Research* Band 4, 2004, S. 9–21.

[Pala06] Otto Palandt. *Bürgerliches Gesetzbuch*. Nr. 7 der Reihe Beck'sche Kurz-Kommentare. C.H. Beck, München. 65. Auflage, 2006.

[Pazz99] Michael J. Pazzani. A Framework for Collaborative, Content-Based and Demographic Filtering. *Artificial Intelligence Review* Band 13, 1999, S. 393–408.

[PGES05] Johan A. Pouwelse, Pawel Garbacki, Dick H. J. Epema und Henk J. Sips. The Bittorrent P2P File-sharing System: Measurements and Analysis. In Miguel Castro und Robbert van Renesse (Hrsg.), *Peer-to-Peer Systems IV, 4th International Workshop, IPTPS 2005, Ithaca, NY, USA, February 24-25, 2005, Revised Selected Papers*, Band 3640 der Reihe *Lecture Notes in Computer Science*, Berlin, Heidelberg, 2005. Springer, S. 205–215.

[PrBe05] Christian Prehofer und Christian Bettstetter. Self-Organization in Communication Networks: Principles and Design Paradigms. *IEEE Communications Magazine* 43(7), 2005, S. 78–85.

[PSWS05] J. A. Pouwelse, M. van Slobbe, J. Wang und H. J. Sips. P2P-based PVR Recommendation using Friends, Taste Buddies and Superpeers. In *Beyond Personalization 2005, Workshop on the Next Stage of Recommender Systems Research, San Diego, California, January 9, 2005*, 2005.

[Raab03] Oliver Raabe. Die rechtliche Einordnung zweier Web-Anonymisierungsdienste. *Datenschutz und Datensicherheit* 27(3), 2003, S. 134–138.

[RaDH07] Oliver Raabe, Jochen Dinger und Hannes Hartenstein. Telekommunikationsdienste in Next-Generation-Networks am Beispiel von Peer-to-Peer-Overlay-Systemen. *Kommunikation und Recht*, März 2007. Beilage 1/2007.

[RCNS06] Daniel Rolli, Michael Conrad, Dirk Neumann und Christoph Sorge. Distributed Ascending Proxy Auction: A Cryptographic Approach. *Wirtschaftsinformatik* 48(1), Februar 2006, S. 7–15.

[RePo87] J.K. Reynolds und J. Postel. Official Internet protocols. RFC 1011, Mai 1987.

[ReRu98] Michael K. Reiter und Aviel D. Rubin. Crowds: Anonymity for Web Transactions. *ACM Transactions on Information and System Security* 1(1), 1998, S. 66–92.

[ReSc01] Ulrich Reber und Mirjam Schorr. Peer-to-Peer-Kommunikationsplattformen und deren Freistellung von der urheberrechtlichen Verantwortlichkeit. *Zeitschrift für Urheber- und Medienrecht* 45(8–9), 2001, S. 672–685.

[ReSG98] Michael G. Reed, Paul F. Syverson und David M. Goldschlag. Anonymous Connections and Onion Routing. *IEEE Journal on Selected Areas in Communications* 16(4), May 1998, S. 482–494.

[ReSR04] Kurt Rebmann, Franz Jürgen Säcker und Roland Rixecker (Hrsg.). *Münchener Kommentar zum Bürgerlichen Gesetzbuch*, Band 5: Schuldrecht · Besonderer Teil III. C.H. Beck, München. 4. Auflage, 2004.

[ReVa97] Paul Resnick und Hal R. Varian. Recommender systems. *Communications of the ACM* 40(3), 1997, S. 56–58.

[ReVa03] Patrick Reynolds und Amin Vahdat. Efficient Peer-to-Peer Keyword Searching. In Markus Endler und Douglas C. Schmidt (Hrsg.), *Middleware 2003, ACM/IFIP/USENIX International Middleware Conference, Rio de Janeiro, Brazil, June 16-20, 2003, Proceedings*, Band 2672 der Reihe *Lecture Notes in Computer Science*. Springer, 2003, S. 21–40.

[ReZe02] Paul Resnick und Richard Zeckhauser. Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. In M.R. Baye (Hrsg.), *The Economics of The Internet and E-Commerce*, Band 11 der Reihe *Advances in Applied Microeconomics*. Elsevier, 2002.

[RFC1034] Paul V. Mockapetris. Domain names - concepts and facilities. RFC 1034 (Standard), November 1987. Updated by RFCs 1101, 1183, 1348, 1876, 1982, 2065, 2181, 2308, 2535, 4033, 4034, 4035, 4343.

[RFC1036] Mark R. Horton und Rick Adams. Standard for interchange of USENET messages. RFC 1036, Dezember 1987.

[RFC2440] Jon Callas, Lutz Donnerhacke, Hal Finney und Rodney Thayer. OpenPGP Message Format. RFC 2440 (Proposed Standard), November 1998.

[RFC2616] Roy T. Fielding, Jim Gettys, Jeffrey C. Mogul, Henrik Frystyk Nielsen, Larry Masinter, Paul Leach und Tim Berners-Lee. Hypertext Transfer Protocol – HTTP/1.1. RFC 2616 (Draft Standard), Juni 1999. Updated by RFC 2817.

[RFC3022] P. Srisuresh und K. Egevang. Traditional IP Network Address Translator (Traditional NAT). RFC 3022 (Informational), Januar 2001.

[RFC3972] Tuomas Aura. Cryptographically Generated Addresses (CGA). RFC 3972 (Proposed Standard), März 2005.

[RFHK+01] Sylvia Ratnasamy, Paul Francis, Mark Handley, Richard Karp und Scott Schenker. A scalable content-addressable network. In *Proceedings of the ACM SIGCOMM 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, August 27-31, 2001, San Diego, CA, USA*. ACM Press, New York, 2001, S. 161–172.

[RGRK04] Sean C. Rhea, Dennis Geels, Timothy Roscoe und John Kubiatowicz. Handling Churn in a DHT. In *Proceedings of the General Track: 2004 USENIX Annual Technical Conference, June 27 - July 2, 2004, Boston Marriott Copley Place, Boston, MA, USA*, 2004, S. 127–140.

[Ripe01] Matei Ripeanu. Peer-to-Peer Architecture Case Study: Gnutella Network. In *P2P '01: Proceedings of the First International Conference on Peer-to-Peer Computing (P2P'01), Linköping, 27-29 August 2001*, Washington, DC, USA, 2001. IEEE Computer Society, S. 99–100.

[RiSA78] Ronald Rivest, Adi Shamir und Leonard Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM* 21(2), 1978, S. 120–126.

[RiSW96] Ronald L. Rivest, Adir Shamir und David A. Wagner. Time-lock Puzzles and Timed-release Crypto. Technischer Bericht TR-684, Massachusetts Institute of Technology, Cambridge, MA, USA, 1996.

[Rive92] Ronald Rivest. The MD5 Message-Digest Algorithm . RFC 1321 (Informational), April 1992.

[RKMG+01] Naren Ramakrishnan, Benjamin J. Keller, Batul J. Mirza, Ananth Y. Grama und George Karypis. Privacy risks in recommender systems. *IEEE Internet Computing* 5(6), 2001, S. 54–63.

[Rüßm90] Helmut Rüßmann. *Rechtsdogmatik und praktische Vernunft: Symposion zum 80. Geburtstag von Franz Wieacker*, Kapitel Möglichkeit und Grenzen der Gesetzesbindung, S. 35–56. Vandenhoeck und Ruprecht, Göttingen. 1990.

[RoBG03] Alexander Roßnagel, Jürgen Banzhaf und Rüdiger Grimm. *Datenschutz im Electronic Commerce*. Nr. 18 der Reihe Schriftenreihe Kommunikation und Recht. Verlag Recht und Wirtschaft. 2003.

[Roßn99] Alexander Roßnagel (Hrsg.). *Recht der Multimedia-Dienste: Kommentar zum IuKDG und zum MDStV*. C.H. Beck, München. Loseblattsammlung mit Ergänzungslieferungen auf dem Stand April 2005, 1999.

[Roßn00] Alexander Roßnagel. Recht der Multimediadienste 1998/1999. *Neue Zeitschrift für Verwaltungsrecht* (6), 2000, S. 622–633.

[Roßn03] Alexander Roßnagel (Hrsg.). *Handbuch Datenschutzrecht: Die neuen Grundlagen für Wirtschaft und Verwaltung*. C.H. Beck, München. 2003.

[Roßn07] Alexander Roßnagel. Personalisierung in der E-Welt: Aus dem Blickwinkel der informationellen Selbstbestimmung gesehen. *Wirtschaftsinformatik* 49(1), 2007, S. 8–15.

[RoSc00] Alexander Roßnagel und Philip Scholz. Datenschutz durch Anonymität und Pseudonymität – Rechtsfolgen der Verwendung anonymer und pseudonymer Daten. *Multimedia und Recht* 3(12), 2000, S. 721–731.

[RStV07] Begründung zum Neunten Staatsvertrag zur Änderung rundfunkrechtlicher Staatsverträge, 2007.

[RZFK00] Paul Resnick, Richard Zeckhauser, Eric Friedman und Ko Kuwabara. Reputation systems. *Communications of the ACM* 43(12), 2000, S. 45–48.

[Scan06] Camillo Scandura. Entwurf und Implementierung eines P2P-basierten Empfehlungsdienstes. Studienarbeit, Institut für Telematik, Universität Karlsruhe (TH), 2006.

[SCDR04] Atul Singh, Miguel Castro, Peter Druschel und Antony Rowstron. Defending against eclipse attacks on overlay networks. In *EW11: Proceedings of the 11th ACM SIGOPS European workshop: beyond the PC, September 19-22, 2004, Leuven, Belgium*, New York, NY, USA, 2004. ACM Press, New York, S. 21–26.

[SCFY96] Ravi S. Sandhu, Edward J. Coyne, Hal L. Feinstein und Charles E. Youman. Role-Based Access Control Models. *IEEE Computer* 29(2), 1996, S. 38–47.

[Scha02] Peter Schaar. *Datenschutz im Internet: Die Grundlagen*. C.H. Beck, München. 2002.

[Schl98] Rainer Schlittgen. *Einführung in die Statistik*. Oldenbourg, München, Wien. 8. Auflage, 1998.

[Schl04] Stefan Schleipfer. Das 3-Schichten-Modell des Multimediadatenschutzrechts. *Datenschutz und Datensicherheit* 28(12), 2004, S. 727–733.

[Schm00] Peter Schmitz. *TDDSG und das Recht auf informationelle Selbstbestimmung*. Nr. 12 der Reihe Schriftenreihe Information und Recht. C.H. Beck, München. 2000.

[Schm05] Nora Schmidt-Keßeler. Die Pflichtangaben für die Internetseite des Steuerberaters nach dem Teledienstgesetz (TDG). *Deutsches Steuerrecht* 43(50), 2005, S. 2143–2144.

[Schn99] Bruce Schneier. Attack Trees: Modeling Security Threats. *Dr. Dobb's Journal* 24(12), 1999, S. 21–29.

[Schw03a] Peter Schwacke. *Juristische Methodik: Mit Technik der Fallbearbeitung*, Band 3 der Reihe *Verwaltung in Praxis und Wissenschaft*. Deutscher Gemeindeverlag und Verlag W. Kohlhammer, Stuttgart. 4. Auflage, 2003.

[Schw03b] Torsten Schwarze. Das Ende des Schreckens? – Beschränkung der punitive damages durch den US-Supreme Court. *Neue Zeitschrift für Gesellschaftsrecht* (17), 2003, S. 804–807.

[Schw06] Matthias Christoph Schwenke. *Individualisierung und Datenschutz: Rechtskonformer Umgang mit personenbezogenen Daten im Kontext der Individualisierung*. DuD-Fachbeiträge. Deutscher Universitätsverlag, Wiesbaden. 2006.

[ScLa05] Florian Schmitz und Stefan Laun. Die Haftung kommerzieller Meinungsportale im Internet. *Multimedia und Recht* 8(4), 2005, S. 208–213.

[ScLS06] Adolf Schönke, Theodor Lenckner und Horst Schröder. *Strafgesetzbuch*. C.H. Beck, München. 27. Auflage, 2006.

[ScOt05] Hans-Bernd Schäfer und Claus Ott. *Lehrbuch der ökonomischen Analyse des Zivilrechts*. Springer, Berlin, Heidelberg. 4. Auflage, 2005.

[Serj02] Andrei Serjantov. Anonymizing Censorship Resistant Systems. In *Peer-to-Peer Systems: First International Workshop, IPTPS 2002, Cambridge, MA, USA, March 7-8, 2002. Revised Papers*, Band 2429 der Reihe *Lecture Notes in Computer Science*, Berlin, Heidelberg, 2002. Springer, S. 111–120.

[SHA195] Secure hash standard. *Federal Information Processing Standards Publication* (180-1), 1995.

[Sieb99] Ulrich Sieber. Die rechtliche Verantwortlichkeit im Internet – Grundlagen, Ziele und Auslegung von § 5 TDG und § 5 MDStV. *Beilage zu Multimedia und Recht* 2(2), 1999.

[SiMo02] Emil Sit und Robert Morris. Security Considerations for Peer-to-Peer Distributed Hash Tables. In Peter Druschel, M. Frans Kaashoek und Antony I. T. Rowstron (Hrsg.), *Peer-to-Peer Systems: First International Workshop, IPTPS 2002, Cambridge, MA, USA, March 7-8, 2002. Revised Papers*, Band 2429 der Reihe *Lecture Notes in Computer Science*, Berlin / Heidelberg, 2002. Springer, S. 261–269.

[SKKR00] Badrul Sarwar, George Karypis, Joseph Konstan und John Riedl. Analysis of recommendation algorithms for e-commerce. In *EC '00: Proceedings of the 2nd ACM Conference on Electronic Commerce (EC-00), October 17-20, 2000, Minneapolis, MN, USA*. ACM Press, New York, 2000, S. 158–167.

[SMKK+01] Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek und Hari Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In *Proceedings of the ACM SIGCOMM 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, August 27-31, 2001, San Diego, CA, USA*. ACM Press, New York, 2001, S. 149–160.

[SoDZ06] Christoph Sorge, Thomas Dreier und Martina Zitterbart. *Information management and market engineering*, Kapitel Trust and the Law, S. 45–56. Universitätsverlag Karlsruhe, Karlsruhe. September 2006.

[Sorg05] Christoph Sorge. *Softwareagenten: Vertragsschluss, Vertragsstrafe, Reugeld*. Nr. 2 der Reihe Schriften des Zentrums für angewandte Rechtswissenschaft. Universitätsverlag Karlsruhe. Zugleich: Diplomarbeit, Universität Karlsruhe (TH), 2005.

[Sorg07a] Christoph Sorge. A Chord-based Recommender System. In *32nd Annual IEEE Conference on Local Computer Networks (LCN 2007), 15-18 October 2007, Clontarf Castle, Dublin, Ireland, Proceedings*, 2007, S. 157–164.

[Sorg07b] Christoph Sorge. Datenschutz in P2P-basierten Systemen: Peer-to-Peer-Netze jenseits des Filesharing. *Datenschutz und Datensicherheit* 31(2), 2007, S. 102–106.

[Sorg07c] Christoph Sorge. Rechtliche Einordnung Peer-to-Peer-basierter Dienste. In *IT-Recht (Datenschutz, Urheberrecht, Telekommunikationsrecht): Tagungsband des 10. Internationalen Rechtsinformatik-Symposions IRIS 2007*, Stuttgart, 2007. Richard Boorberg, S. 336–344.

[SoZi06] Christoph Sorge und Martina Zitterbart. A Reputation-Based System for Confidentiality Modeling in Peer-To-Peer Networks. In *Trust Management: Fourth International Conference, iTrust 2006, Pisa, Italy, May 16-19, 2006, Proceedings*, Band 3986 der Reihe *Lecture Notes in Computer Science*, Berlin, Heidelberg, 2006. Springer, S. 367–381.

[SpLe05] Gerald Spindler und Matthias Leistner. Die Verantwortlichkeit für Urheberrechtsverletzungen im Internet – Neue Entwicklungen in Deutschland und in den USA. *Gewerblicher Rechtsschutz und Urheberrecht: Internationaler Teil* 54(10), 2005, S. 773–796.

[SpSG04] Gerald Spindler, Peter Schmitz und Ivo Geis. *TDG: Teledienstegesetz, Teledienstedatenschutzgesetz, Signaturgesetz*. C.H. Beck, München. 2004.

[Stad05] Thomas Stadler. *Haftung für Informationen im Internet*, Band 5 der Reihe *Electronic Commerce und Recht*. Erich Schmidt Verlag, Berlin. 2. Auflage, 2005.

[Star05] Christian Starck (Hrsg.). *Kommentar zum Grundgesetz*. Franz Vahlen. Begründet von Hermann von Mangoldt, fortgeführt von Friedrich Klein, 5. Auflage, 2005.

[Stic04] Barbara Stickelbrock. „Impressumspflicht“ im Internet – eine kritische Analyse der neueren Rechtsprechung zur Anbieterkennzeichnung nach § 6 TDG. *Gewerblicher Rechtsschutz und Urheberrecht* 106(2), 2004, S. 111–117.

[StWe04] Ralf Steinmetz und Klaus Wehrle. Peer-to-Peer-Networking & -Computing. *Informatik Spektrum* 27(1), Februar 2004, S. 51–54.

[SyGR97] Paul F. Syverson, David M. Goldschlag und Michael G. Reed. Anonymous Connections and Onion Routing. In *IEEE Symposium on Security and Privacy*, 1997, S. 44–54.

[Szto01] Piotr Sztompka. *International Encyclopedia of the Social and Behavioral Sciences*, Band 23, Kapitel Trust: Cultural Concerns, S. 15913–15917. Elsevier, Amsterdam, Paris, New York. 2001.

[TaGV02] Jörg Tauss, Monika Griefahn, Ute Vogt et al. Umfassende Modernisierung des Datenschutzrechtes voranbringen, 2002. Antrag der Fraktion der SPD und der Fraktion Bündnis 90/Die Grünen, Deutscher Bundestag, Drucksache 14/9709.

[Tett99] Alexander Tettenborn. Die Evaluierung des IuKDG – Erfahrungen, Erkenntnisse und Schlußfolgerungen. *Multimedia und Recht* 2(9), 1999, S. 516–522.

[TrWa02] Wade Trappe und Lawrence C. Washington. *Introduction to cryptography: with coding theory*. Prentice Hall, New Jersey. 2002.

[Ulbr04] Johannes Ulbricht. Tücken im Schutz für Kopierschutz. *Computer und Recht* 20(9), 2004, S. 674–679.

[vSta99] Julius von Staudinger. *Kommentar zum Bürgerlichen Gesetzbuch mit Einführungsgesetz und Nebengesetzen*, Band §§ 823-825. Sellier-de Gruyter, Berlin. Bearbeitet von Johannes Hager, 1999.

[WaBr90] Samuel D. Warren und Louis D. Brandeis. The Right to Privacy. *Harvard Law Review* 4(5), 1890, S. 193–220.

[Wald98] Arthur Waldenberger. Teledienste, Mediendienste und die „Verantwortlichkeit“ ihrer Anbieter. *Multimedia und Recht* 1(3), 1998, S. 124–129.

[Wall03] Dan S. Wallach. A Survey of Peer-to-Peer Security Issues. In Mitsuhiro Okada, Benjamin C. Pierce, Andre Scedrov, Hideyuki Tokuda und Akinori Yonezawa (Hrsg.), *Software Security – Theories and Systems, Mext-NSF-JSPS, International Symposium, ISSS 2002, Tokyo, Japan, November 8-10, 2002, Revised Papers*, Band 2609 der Reihe *Lecture Notes in Computer Science*, Berlin, Heidelberg, 2003. Springer, S. 42–57.

[WaSi05] Kevin Walsh und Emin Gün Sirer. Fighting peer-to-peer SPAM and decoys with object reputation. In Eric Friedman und Emin Gün Sirer (Hrsg.), *Proceedings of the ACM SIGCOMM Third Workshop on the Economics of Peer-to-Peer Systems, P2PECON '05 : Monday, August 22nd 2005, Philadelphia, Pennsylvania, USA*. ACM Press, New York, 2005, S. 138–143. Enthalten in SIGCOMM 2005: proceedings of ACM SIGCOMM 2005 workshops; E-Wind 2005, P2Pecon '05, MineNet '05, WDTN '05 ; Conference on Computer Communications, August 22nd & 26th, 2005, Philadelphia, PA.

[WaYu05] Xiaoyun Wang und Hongbo Yu. How to break MD5 and other Hash Functions. In *Advances in Crptology – Eurocypt 2005, 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Den-*

mark, May 22–26, 2005, Proceedings, Band 3494 der Reihe *Lecture Notes in Computer Science*, Berlin, Heidelberg, 2005. Springer, S. 19–35.

[Webe02] Jan Weber. Der Adressatenkreis der Verpflichtung zur Anbieterkennung im Internet nach der Neufassung des Teledienstegesetzes. *JurPC: Internet-Zeitschrift für Rechtsinformatik und Informationsrecht* (76), 2002. JurPC Web-Dokument 76/2002, `http://www.jurpc.de/aufsatz/20020076.htm`, abgerufen am 6.9.2006.

[Wein97] Jürgen Weinknecht. Kommentierung zu § 2 TDG. *Online Journal Recht*, 1997. `http://www.weinknecht.de/ojr/gesetze/TDG.htm`. Veröffentlichungsdatum unsicher. Abgerufen am 30.6.2006.

[Weis99] Gerhard Weiss. *Multiagent systems: a modern approach to distributed artificial intelligence*. The MIT Press, Cambridge, London. 1999.

[Wenz05] Frauke Wenzl. *Musiktauschbörsen im Internet: Haftung und Rechtsschutz nach deutschem und amerikanischem Urheberrecht*. Nr. 196 der Reihe Wirtschaftsrecht und Wirtschaftspolitik. Nomos, Baden-Baden. 2005.

[West67] Alan F Westin. *Privacy and freedom*. Atheneum, New York. 1967.

[Will93] Oliver E. Williamson. Calculativeness, trust, and economic organization. *The Journal of Law & Economics* Band 36, April 1993, S. 453–486.

[WoGe05] Hans H. Wohlgemuth und Jürgen Gerloff. *Datenschutzrecht: Eine Einführung mit praktischen Fällen*. Luchterhand, München/Unterschleißheim. 3. Auflage, 2005.

[Woit03] Thomas Woitke. Das „Wie“ der Anbieterkennzeichnung gemäß § 6 TDG. *Neue Juristische Wochenschrift* 56(12), 2003, S. 871–873.

[Wool02] Michael Wooldridge. *An introduction to MultiAgent Systems*. Wiley, Chichester. 2002.

[WPLR06] Jun Wang, Johan Pouwelse, Reginald Lagendijk und Marcel R. J. Reinders. Distributed Collaborative Filtering for Peer-to-Peer File Sharing Systems. In *Proceedings of the 21st ACM Symposium on Applied Computing (SAC'06), April 2006, Dijon, France*. ACM Press, New York, 2006.

[Wätj03] Dietmar Wätjen. *Kryptographie: Grundlagen, Algorithmen, Protokolle*. Spektrum Akademischer Verlag, Heidelberg, Berlin. 2003.

[Wund07] Lars Wunderlich. Vertraulichkeitsmodellierung in Peer-to-Peer-Netzen. Studienarbeit, Institut für Telematik, Universität Karlsruhe (TH), 2007.

[WöWü05] Thomas Wölfl und Sven Wünschmann. Public-Key-Infrastrukturen in einer Peer-to-Peer-Umgebung. In Armin B. Cremers, Rainer Manthey, Peter Martini und Volker Steinhage (Hrsg.), *INFORMATIK 2005 - Informatik LIVE! Band 2, Beiträge der 35. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Bonn, 19. bis 22. September 2005 (2)*, 2005, S. 643–647.

[XHYS04] Bo Xie, Peng Han, Fan Yang und Ruimin Shen. An Efficient Neighbor Searching Scheme of Distributed Collaborative Filtering on P2P Overlay Network. In Fernando Galindo, Makoto Takizawa und Roland Traunmüller (Hrsg.), *Database and Expert Systems Applications, 15th International Conference, DEXA 2004 Zaragoza, Spain, August 30-September 3, 2004, Proceedings*, Band 3180 der Reihe *Lecture Notes in Computer Science*, Berlin, Heidelberg, 2004. Springer, S. 141–150.

[XLYX+05] Gui-Rong Xue, Chenxi Lin, Qiang Yang, WenSi Xi, Hua-Jun Zeng, Yong Yu und Zheng Chen. Scalable collaborative filtering using cluster-based smoothing. In Ricardo A. Baeza-Yates, Nivio Ziviani, Gary Marchionini, Alistair Moffat und John Tait (Hrsg.), *SIGIR 2005: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Salvador, Brazil, August 15-19, 2005*. ACM Press, New York, 2005, S. 114–121.

[YaYa94] Toshio Yamagishi und Midori Yamagishi. Trust and Commitment in the United States and Japan. *Motivation and Emotion* 18(2), 1994, S. 129–166.

[ZaMP98] Akbar Zaheer, Bill McEvily und Vincenzo Perrone. Does Trust Matter? Exploring the Effects of Interorganizational and Interpersonal Trust on Performance. *Organization Science* 9(2), 1998, S. 141–159.

[Zieg05] Cai-Nicolas Ziegler. *Towards Decentralized Recommender Systems*. Dissertation, Albert-Ludwigs-Universität, Freiburg, 2005.

[Zimm80] Hubert Zimmermann. OSI Reference Model – The ISO Model of Architecture for Open Systems Interconnection. *IEEE Transactions on Communications* 28(4), April 1980, S. 425–432.

[ZMKL05] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan und Georg Lausen. Improving recommendation lists through topic diversification. In *WWW '05: Proceedings of the 14th international conference on World Wide Web (Chiba, Japan, May 10–14, 2005)*, New York, NY, USA, 2005. ACM Press, S. 22–32.