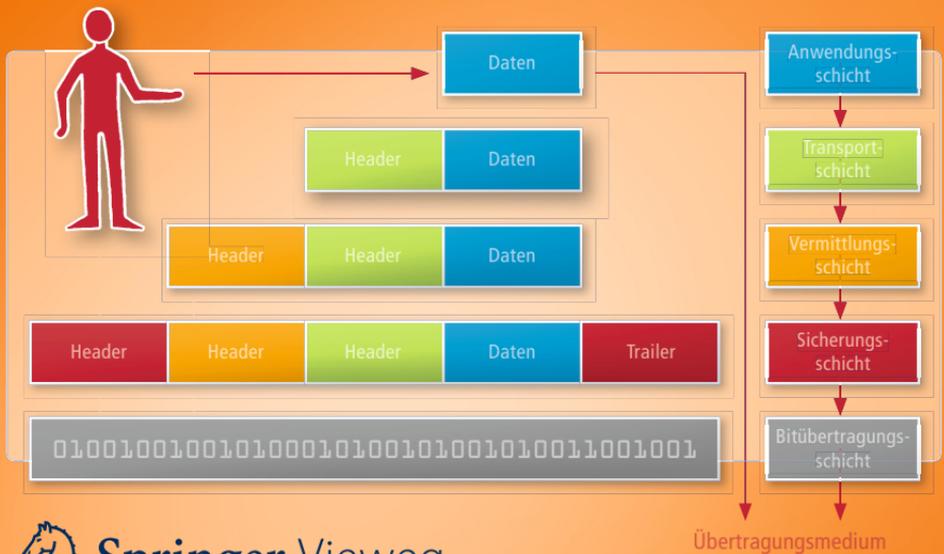


Christian Baun

Computernetze kompakt

3. Auflage



Springer Vieweg

IT kompakt

Werke der „kompakt-Reihe“ zu wichtigen Konzepten und Technologien der IT-Branche:

- ermöglichen einen raschen Einstieg,
- bieten einen fundierten Überblick,
- sind praxisorientiert, aktuell und immer ihren Preis wert.

Weitere Titel der Reihe siehe: <http://www.springer.com/series/8297>.

Christian Baun

Computernetze kompakt

3. Auflage



Springer Viewweg

Christian Baun
Fachbereich Informatik
Fachhochschule Frankfurt am Main
Frankfurt, Deutschland

ISBN 978-3-662-46931-6

ISBN 978-3-662-46932-3 (eBook)

DOI 10.1007/978-3-662-46932-3

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

Springer Vieweg

© Springer-Verlag Berlin Heidelberg 2012, 2013, 2015

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

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen.

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier.

Springer Berlin Heidelberg ist Teil der Fachverlagsgruppe Springer Science+Business Media

(www.springer.com)

Vorwort

Vorwort zur 3. Auflage

Erneut erscheint nach wenig mehr als einem Jahr eine neue Auflage. Diese enthält im Vergleich zur 2. Auflage zahlreiche Korrekturen. Verbessert wurden unter anderem die Abbildungen zu Broadcast-Domänen und Kollisionsdomänen und die Unterscheidung zwischen Router und Layer-3-Switch.

Neu hinzugekommen ist Kapitel 11, das eine Übersicht über wichtige Kommandozeilenwerkzeuge zur Netzwerkkonfiguration und Analyse von Netzwerkproblemen enthält. Der Aufbau des Kapitels orientiert sich an den Schichten im hybriden Referenzmodell und damit auch am Aufbau dieses Buches.

An dieser Stelle möchte ich Kristian Kraljic, Tobias Kurze, Michael Stapelberg, Torsten Wiens und ganz besonders Katrin Baun für das Korrekturlesen des neuen Kapitels danken.

Frankfurt am Main

Prof. Dr. Christian Baun
Februar 2015

Vorwort zur 2. Auflage

Aufgrund der positiven Rückmeldungen und Verbesserungsvorschläge haben ich nach nur knapp einem Jahr seit Erscheinen der 1. Auflage einige Teile des Buches überarbeitet und erweitert.

Neu hinzugekommen ist Abschnitt 3.10, der eine Beschreibung der Themen Bandbreite und Latenz und des damit verbundenen Bandbreite-Verzögerung-Produkts enthält.

Umfangreich erweitert wurde das Thema Rahmen bei WLAN in Abschnitt 6.4. Neu hinzugekommen ist ebenfalls Abschnitt 6.5, der die Maximum Transmission Unit (MTU) beschreibt.

Die sehr knappe Beschreibung der verschiedenen Routing-Protokolle ist einer detaillierteren Betrachtung der beiden Routing-Protokolle RIP und OSPF in den Abschnitten 7.4 und 7.5 gewichen. Neu hinzugekommen ist eine ausführliche Beschreibung des *Internetworking*, also der Kommunikation über verschiedene physische Netze in Abschnitt 7.6.

Die Beschreibung des Transportprotokolls TCP in Abschnitt 8.4 wurde unter anderem um die Themen Flusskontrolle und Überlastkontrolle erweitert.

Überarbeitet wurden auch die Abschnitte 9.1 und 9.2, in denen die Anwendungsprotokolle DNS und DHCP beschrieben sind.

Komplett neu hinzugekommen ist Kapitel 10, das eine Einführung in die Netzwerkvirtualisierung enthält.

Frankfurt am Main

Prof. Dr. Christian Baum
September 2013

Vorwort zur 1. Auflage

Computernetze sind ein dauerhaft aktuelles Thema der Informatik. Waren in der Vergangenheit die kabelgebundene Vernetzung von PCs und Workstations sowie das globale Internet dominierende Themen, sind es heute zunehmend mobile Systeme und Funknetze.

Dieses kompakte Werk über das breite Thema Computernetze wurde mit dem Ziel geschrieben, dem Leser einen Überblick über die wichtigsten Vernetzungstechnologien, Netzwerkgeräte, Protokolle und Übertragungsmedien zu verschaffen und so das Verständnis dafür zu wecken, wie Computernetze funktionieren.

Für das Korrekturlesen danke ich Maximilian Hoecker, Kristian Kraljic, Michael Stapelberg und Torsten Wiens.

An dieser Stelle möchte ich ganz besonders meiner Familie danken, die mich auch bei diesem Werk so viel unterstützt hat.

Mannheim

Dr. Christian Baun

Juli 2012

Inhaltsverzeichnis

1	Einleitung	1
2	Grundlagen der Informationstechnik	3
2.1	Bit	3
2.2	Repräsentation von Zahlen	4
2.2.1	Dezimalsystem	5
2.2.2	Dualsystem	5
2.2.3	Oktalsystem	6
2.2.4	Hexadezimalsystem	6
2.3	Datei- und Speichergrößen	8
2.4	Informationsdarstellung	9
2.4.1	ASCII-Kodierung	10
2.4.2	Unicode	12
2.4.3	Darstellung von Zeichenketten	13
3	Grundlagen der Computervernetzung	15
3.1	Entwicklung der Computernetze	15
3.2	Komponenten von Computernetzen	17
3.3	Räumliche Ausdehnung von Computernetzen	17
3.4	Datenübertragung	18
3.4.1	Serielle und parallele Übertragung	19
3.4.2	Synchrone und asynchrone Übertragung	19
3.4.3	Richtungsabhängigkeit der Übertragung	20
3.5	Geräte in Computernetzen	21

3.6	Topologien von Computernetzen	23
3.6.1	Bus-Topologie	23
3.6.2	Ring-Topologie	25
3.6.3	Stern-Topologie	25
3.6.4	Maschen-Topologie	26
3.6.5	Baum-Topologie	26
3.6.6	Zellen-Topologie	27
3.7	Frequenz und Datensignal	27
3.8	Fourierreihe und Bandbreite	28
3.9	Bitrate und Baudrate	28
3.10	Bandbreite und Latenz	29
3.10.1	Bandbreite-Verzögerung-Produkt	30
3.11	Zugriffsverfahren	30
3.11.1	Deterministisches Zugriffsverfahren	31
3.11.2	Nicht-deterministisches Zugriffsverfahren	31
3.12	Kollisionsdomäne (Kollisionsgemeinschaft)	32
4	Protokolle und Protokollschichten	33
4.1	TCP/IP-Referenzmodell	34
4.2	Hybrides Referenzmodell	35
4.2.1	Bitübertragungsschicht	35
4.2.2	Sicherungsschicht	36
4.2.3	Vermittlungsschicht	37
4.2.4	Transportschicht	37
4.2.5	Anwendungsschicht	38
4.3	Ablauf der Kommunikation	39
4.4	OSI-Referenzmodell	40
4.4.1	Sitzungsschicht	40
4.4.2	Darstellungsschicht	41
4.5	Fazit zu den Referenzmodellen	41
5	Bitübertragungsschicht	43
5.1	Vernetzungstechnologien	43
5.1.1	Ethernet	43
5.1.2	Token Ring	45
5.1.3	Wireless Local Area Network (WLAN)	46
5.1.4	Bluetooth	55

5.2	Übertragungsmedien	58
5.2.1	Koaxialkabel	58
5.2.2	Twisted-Pair-Kabel	60
5.2.3	Lichtwellenleiter	64
5.3	Strukturierte Verkabelung	65
5.4	Geräte der Bitübertragungsschicht	66
5.4.1	Auswirkungen von Repeatern und Hubs auf die Kollisionsdomäne	68
5.5	Kodierung von Daten in Netzwerken	69
5.5.1	Non-Return to Zero (NRZ)	71
5.5.2	Non-Return to Zero Invert (NRZI)	72
5.5.3	Multilevel Transmission Encoding – 3 Levels (MLT-3)	73
5.5.4	Return-to-Zero (RZ)	73
5.5.5	Unipolares RZ	74
5.5.6	Alternate Mark Inversion (AMI)	75
5.5.7	Bipolar with 8 Zeros Substitution (B8ZS)	75
5.5.8	Manchester	76
5.5.9	Manchester II	77
5.5.10	Differentielle Manchesterkodierung	77
5.6	Nutzdaten mit Blockcodes verbessern	78
5.6.1	4B5B	79
5.6.2	5B6B	79
5.6.3	8B10B-Kodierung	81
5.7	Weitere Leitungscodes	82
5.7.1	8B6T-Kodierung	83
6	Sicherungsschicht	85
6.1	Geräte der Sicherungsschicht	85
6.1.1	Lernende Bridges	87
6.1.2	Kreise auf der Sicherungsschicht	88
6.1.3	Spanning Tree Protocol (STP)	88
6.1.4	Auswirkungen von Bridges und Layer-2-Switches auf die Kollisionsdomäne	92
6.2	Adressierung in der Sicherungsschicht	92
6.2.1	Format der MAC-Adressen	93

	6.2.2	Eindeutigkeit von MAC-Adressen	93
	6.2.3	Sicherheit von MAC-Adressen	94
6.3		Rahmen abgrenzen	95
	6.3.1	Längenangabe im Header	95
	6.3.2	Zeichenstopfen	96
	6.3.3	Bitstopfen	96
	6.3.4	Verstöße gegen Regeln des Leitungscodes	97
6.4		Rahmenformate aktueller Computernetze	98
	6.4.1	Rahmen bei Ethernet	98
	6.4.2	Rahmen bei WLAN	99
	6.4.3	Spezielle Rahmen bei WLAN	102
6.5		Maximum Transmission Unit (MTU)	103
6.6		Fehlererkennung	104
	6.6.1	Zweidimensionale Parität	104
	6.6.2	Zyklische Redundanzprüfung	104
6.7		Medienzugriffsverfahren	106
	6.7.1	Medienzugriffsverfahren bei Ethernet	107
	6.7.2	Medienzugriffsverfahren bei WLAN	110
6.8		Adressauflösung mit dem Address Resolution Protocol	115
7		Vermittlungsschicht	117
	7.1	Geräte der Vermittlungsschicht	118
		7.1.1 Auswirkungen von Routern auf die Kollisionsdomäne	118
		7.1.2 Broadcast-Domäne (Rundsendedomäne)	118
	7.2	Adressierung in der Vermittlungsschicht	119
		7.2.1 Aufbau von IP-Adressen	121
		7.2.2 Subnetze	122
		7.2.3 Private IP-Adressen	125
		7.2.4 Aufbau von IP-Paketen	125
		7.2.5 Fragmentieren von IP-Paketen	128
	7.3	Weiterleitung und Wegbestimmung	128
	7.4	Routing Information Protocol (RIP)	130
		7.4.1 Count-to-Infinity	130
		7.4.2 Split Horizon	131
		7.4.3 Fazit zu RIP	133

7.5	Open Shortest Path First (OSPF)	136
7.5.1	Routing-Hierarchie mit OSPF	136
7.5.2	Arbeitsweise von OSPF	137
7.5.3	Aufbau von OSPF-Nachrichten	139
7.5.4	Fazit zu OSPF	141
7.6	Netzübergreifende Kommunikation	142
7.7	Diagnose und Fehlermeldungen mit ICMP	144
8	Transportschicht	147
8.1	Eigenschaften von Transportprotokollen	147
8.2	Adressierung in der Transportschicht	148
8.3	User Datagram Protocol (UDP)	149
8.3.1	Aufbau von UDP-Segmenten	150
8.4	Transmission Control Protocol (TCP)	151
8.4.1	Aufbau von TCP-Segmenten	152
8.4.2	Arbeitsweise von TCP	154
8.4.3	Flusskontrolle	158
8.4.4	Überlastkontrolle	161
9	Anwendungsschicht	169
9.1	Domain Name System (DNS)	169
9.1.1	Arbeitsweise des DNS	170
9.1.2	Auflösung eines Domainnamens	171
9.2	Dynamic Host Configuration Protocol (DHCP)	172
9.2.1	Arbeitsweise von DHCP	173
9.2.2	Aufbau von DHCP-Nachrichten	175
9.2.3	DHCP-Relay	176
9.3	Telecommunication Network (Telnet)	177
9.4	Hypertext Transfer Protocol (HTTP)	178
9.5	Simple Mail Transfer Protocol (SMTP)	181
9.6	Post Office Protocol Version 3 (POP3)	182
9.7	File Transfer Protocol (FTP)	182
10	Netzwerkvirtualisierung	185
10.1	Virtual Private Networks (VPN)	185
10.1.1	Technische Arten von VPNs	186
10.2	Virtual Local Area Networks (VLAN)	187

11	Kommandozeilenwerkzeuge	191
11.1	ethtool	194
11.2	mii-tool	194
11.3	iwconfig	195
11.4	iwlist	196
11.5	arp	196
11.6	ifconfig	197
11.7	ip	198
11.8	tcpdump	199
11.9	route	200
11.10	ping	200
11.11	traceroute	201
11.12	netstat	201
11.13	nmap	202
11.14	dhclient	203
11.15	dig	203
11.16	ftp	204
11.17	netperf	204
11.18	nslookup	205
11.19	ssh	205
11.20	telnet	205
Glossar		207
Literatur		213
Sachverzeichnis		215

Dieses Buch will einen Überblick über Computernetze und deren Komponenten schaffen, ohne dabei den Anspruch auf Vollständigkeit zu erheben. Das Ziel ist es, den Leserinnen und Lesern ein grundlegendes Wissen über die Funktionsweise moderner Computernetze und deren Komponenten zu vermitteln. Technische Vorkenntnisse sind dabei nicht erforderlich.

In den Kap. 2 und 3 findet eine Einführung in die Grundlagen der Informationstechnik (IT) und der Computervernetzung statt. Dies ist nötig, um die Thematik Computernetze und den Inhalt dieses Buchs verstehen zu können.

Kapitel 4 beschäftigt sich mit den Grundlagen von Protokollen und deren Einordnung in etablierte Referenzmodelle. Das Kapitel beschreibt auch den Ablauf der Kommunikation und den Weg der Nutzdaten durch die einzelnen Protokollschichten.

In den Kap. 5 bis 9 werden die einzelnen Protokollschichten von der untersten Schicht, der Bitübertragungsschicht, bis zur obersten Schicht, der Anwendungsschicht, behandelt.

In Kap. 10 findet eine Einführung in die Varianten der Netzwerkvirtualisierung statt.

Kap. 11 enthält eine Übersicht über wichtige Kommandozeilenwerkzeuge zur Netzwerkkonfiguration und Analyse von Netzwerkproblemen.

Grundlagen der Informationstechnik 2

Um die Funktionsweise der Computernetze zu verstehen, ist ein grundlegendes Verständnis der Informationstechnik (IT) nötig. Bei diesen Grundlagen handelt es sich um die Möglichkeiten der Informationsdarstellung und Repräsentation von Zahlen, Größenordnungen und die Art und Weise, wie Informationen (speziell Texte) in Rechnern dargestellt werden.

2.1 Bit

Ein *Bit* ist die kleinstmögliche Einheit der Information und jede Information ist an einen Informationsträger gebunden [3]. Ein Informationsträger, der sich in genau einem von zwei Zuständen befinden kann, kann die Datenmenge 1 Bit darstellen. Den Wert eines oder mehrerer Bits nennt man *Zustand*. Ein Bit stellt zwei Zustände dar. Verschiedene Sachverhalte können die Datenmenge 1 Bit darstellen. Beispiele sind:

- Die Stellung eines Schalters mit zwei Zuständen
- Der Schaltzustand eines Transistors
- Das Vorhandensein einer elektrischen Spannung oder Ladung
- Das Vorhandensein einer Magnetisierung
- Der Wert einer Variable mit den logischen Wahrheitswerten.

Benötigt man zur Speicherung einer Information mehr als zwei Zustände, sind Folgen von Bits (*Bitfolgen*) nötig. Mit n Bits kann man 2^n verschiedene Zustände darstellen (siehe Tab. 2.1). Also kann man mit 2 Bits

Tab. 2.1 Die Anzahl der darstellbaren Zustände verdoppelt sich mit jedem zusätzlichen Bit

Bits	Zustände	Bits	Zustände	Bits	Zustände
1	$2^1 = 2$	9	$2^9 = 512$	17	$2^{17} = 131.072$
2	$2^2 = 4$	10	$2^{10} = 1.024$	18	$2^{18} = 262.144$
3	$2^3 = 8$	11	$2^{11} = 2.048$	19	$2^{19} = 524.288$
4	$2^4 = 16$	12	$2^{12} = 4.096$	20	$2^{20} = 1.048.576$
5	$2^5 = 32$	13	$2^{13} = 8.192$	21	$2^{21} = 2.097.152$
6	$2^6 = 64$	14	$2^{14} = 16.384$	22	$2^{22} = 4.194.304$
7	$2^7 = 128$	15	$2^{15} = 32.768$	23	$2^{23} = 8.388.608$
8	$2^8 = 256$	16	$2^{16} = 65.536$	24	$2^{24} = 16.777.216$

$2^2 = 4$ verschiedene Zustände repräsentieren, nämlich 00, 01, 10 und 11. Mit 3 Bits kann man schon $2^3 = 8$ verschiedene Zustände (000, 001, 010, 011, 100, 101, 110 und 111) Zustände repräsentieren. Jedes zusätzliche Bit verdoppelt die Anzahl der möglichen darstellbaren Zustände (Bitfolgen) [3].

2.2 Repräsentation von Zahlen

Zahlen kann man auf unterschiedliche Arten darstellen. Eine Aufgabe der IT ist es, Zahlen aus der *realen Welt* im Computer abzubilden. Wichtig ist dabei die Unterscheidung zwischen *Wert* und *Darstellung*.

In der Mathematik unterscheidet man Zahlen als Elemente verschiedener Wertemengen (natürliche Zahlen, ganze Zahlen, reelle Zahlen, komplexe Zahlen, usw.). Den Wert einer Zahl nennt man auch *abstrakte Zahl* und der Wert ist unabhängig von der Darstellung (zum Beispiel $0,5 = 1/2$).

Operationen eines Rechners werden aber nicht auf Werten, sondern auf Bitfolgen ausgeführt. Darum ist für die IT besonders die Darstellung der Zahlen interessant. Die Darstellung wird vom verwendeten Stellenwertsystem (Positionssystem) bestimmt. Die für die IT wichtigen Stellenwertsysteme sind das Dezimalsystem, das Dualsystem, das Oktalsystem und das Hexadezimalsystem.

2.2.1 Dezimalsystem

Das Dezimalsystem verwendet als Basis die Zahl 10. Jede Ziffer D an der Stelle i hat den Wert $D \times 10^i$. Ein Beispiel ist:

$$2.013 = 2 \times 10^3 + 0 \times 10^2 + 1 \times 10^1 + 3 \times 10^0$$

Computer-Systeme unterscheiden prinzipiell zwischen zwei elektrischen Zuständen. Darum ist aus Sicht der IT als Basis die Zahl 2 und damit das Dualsystem optimal geeignet.

2.2.2 Dualsystem

Das Dualsystem verwendet als Basis die Zahl 2. Zahlen werden nur mit den Ziffern der Werte Null und Eins dargestellt. Zahldarstellungen im Dualsystem heißen Dualzahlen oder Binärzahlen. Alle positiven natürlichen Zahlen inklusive der Null können durch Folgen von Symbolen aus der Menge $\{1, 0\}$ repräsentiert werden. Das niederwertigste Bit (x_0) heißt *Least Significant Bit* (LSB) und das höchstwertigste Bit (x_{n-1}) ist das *Most Significant Bit* (MSB).

Da lange Reihen von Nullen und Einsen für Menschen schnell unübersichtlich werden, verwendet man zur Darstellung von Bitfolgen häufig das Oktalsystem oder das Hexadezimalsystem.

Die Umrechnung der Stellenwertsysteme ist einfach möglich. Zur Verdeutlichung ist das Stellenwertsystem der jeweiligen Zahl in den folgenden Beispielen tiefgestellt beigefügt.

Bei der Umwandlung von Dualzahlen in Dezimalzahlen werden die Ziffern mit ihren Stellenwertigkeiten ausmultipliziert und die Ergebnisse addiert.

$$10100100_2 = 2^7 + 2^5 + 2^2 = 164_{10}$$

Die Umwandlung von Dezimalzahlen in Dualzahlen ist unter anderem mit Hilfe des Restwertalgorithmus (siehe Tab. 2.2) möglich. Dabei wird die Dezimalzahl durch die Basis 2 dividiert und das Ergebnis und der Rest (Wert Null oder Eins) werden notiert. Das Ergebnis der Division wird in der nächsten Runde (Zeile der Tabelle) erneut durch die Basis

Tab. 2.2 Die Dezimalzahl 164_{10} in die Dualzahl 10100100_2 umwandeln

<i>Quotient</i>		<i>Rest</i>
k	$k \text{ DIV } 2$	$k \text{ MODULO } 2$
164	82	$0 = x_1$
82	41	$0 = x_2$
41	20	$1 = x_3$
20	10	$0 = x_4$
10	5	$0 = x_5$
5	2	$1 = x_6$
2	1	$0 = x_7$
1	0	$1 = x_8$

dividiert und erneut werden das Ergebnis und der Rest notiert. Der Algorithmus wird so lange weitergeführt, bis das Ergebnis der Division Null ist.

2.2.3 Oktalsystem

Das Oktalsystem verwendet als Basis die Zahl 8 und kann Gruppen von 3 Bits (Oktade) mit einem Zeichen darstellen.

Bei der Umwandlung von Dualzahlen in Oktalzahlen wird die Bitkette vom niederwertigsten Bit beginnend in Oktaden unterteilt. Jede Oktade entspricht einer Stelle der Oktalzahl.

$$164_{10} = 10|100|100_2 = 244_8$$

Die Umwandlung von Oktalzahlen in Dualzahlen erfolgt analog. Eine Stelle im Oktalsystem entspricht drei Stellen im Dualsystem.

2.2.4 Hexadezimalsystem

Das Hexadezimalsystem verwendet als Basis die Zahl 16. Die Darstellung positiver natürlicher Zahlen erfolgt mit den 16 Ziffernsymbolen aus der Menge $\{0, 1, \dots, 8, 9, A, B, C, D, E, F\}$. Ein Zeichen kann eine Gruppe von 4 Bits (Tetrade oder Nibble) darstellen.

Tab. 2.3 Verschiedene Darstellungen positiver natürlicher Zahlen

Dezimale Darstellung	Binäre Darstellung	Oktale Darstellung	Hexadezimale Darstellung
00	0000	00	0
01	0001	01	1
02	0010	02	2
03	0011	03	3
04	0100	04	4
05	0101	05	5
06	0110	06	6
07	0111	07	7
08	1000	10	8
09	1001	11	9
10	1010	12	A
11	1011	13	B
12	1100	14	C
13	1101	15	D
14	1110	16	E
15	1111	17	F

Bei der Umwandlung von Dualzahlen in Hexadezimalzahlen wird die Bitkette vom niederwertigsten Bit beginnend in Tetraden unterteilt. Jede Tetrade entspricht einer Stelle der Hexadezimalzahl.

$$164_{10} = 1010|0100_2 = A4_{16}$$

Die Umwandlung von Hexadezimalzahlen in Dualzahlen geht analog. Eine Stelle im Hexadezimalsystem entspricht vier Stellen im Dualsystem.

Tabelle 2.3 enthält eine Übersicht der verschiedenen Darstellungen der ersten 16 positiven natürlichen Zahlen im Dezimalsystem, Dualsystem, Oktalsystem und Hexadezimalsystem.

2.3 Datei- und Speichergrößen

Rechner lesen und schreiben aus Geschwindigkeitsgründen meist nicht einzelne Bits, sondern arbeiten mit Bitfolgen, deren Längen Vielfache von Acht sind. Eine Gruppe von 8 Bits nennt man *Byte*. Den Wert eines Bytes kann man entweder durch 8 Bits oder zwei Hexadezimalziffern darstellen.

Eine Datei ist eine beliebig lange Folge von Bytes und enthält inhaltlich zusammengehörende Daten. Alle Informationen (Zahlen, Texte, Musik, Programme, usw.), mit denen ein Rechner arbeiten soll, müssen als Folge von Bytes repräsentiert werden können und als Datei gespeichert werden [3].

Da sich die Größenordnungen der meisten Dateien im Bereich mehrerer Tausend oder Millionen Bytes befinden, gibt es verschiedene Größeneinheiten zur verkürzten Zahlendarstellung. Für Datenspeicher mit binärer Adressierung ergeben sich Speicherkapazitäten von 2^n Byte, also Zweierpotenzen (siehe Tab. 2.4).

Die Maßeinheiten in Tab. 2.4 haben sich für die Größenangabe von Hauptspeicher und Speichermedien in Betriebssystemen eingebürgert. Die Hersteller von Festplatten, CD/DVDs und USB-Speichermedien verwenden zur Berechnung der Kapazität und zur Angabe auf der Verpackung aber lieber Dezimal-Präfixe, also zum Beispiel den Faktor 10^9 anstatt 2^{30} für GB und 10^{12} anstatt 2^{40} für TB. Aus diesem Grund wird zum Beispiel bei einem DVD-Rohling mit einer angegebenen Kapazi-

Tab. 2.4 Datei- und Speichergrößen

Name	Symbol	Bedeutung
Kilobyte	kB	$2^{10} = 1.024$ Bytes
Megabyte	MB	$2^{20} = 1.048.576$ Bytes
Gigabyte	GB	$2^{30} = 1.073.741.824$ Bytes
Terabyte	TB	$2^{40} = 1.099.511.627.776$ Bytes
Petabyte	PB	$2^{50} = 1.125.899.906.842.624$ Bytes
Exabyte	EB	$2^{60} = 1.152.921.504.606.846.976$ Bytes
Zettabyte	ZB	$2^{70} = 1.180.591.620.717.411.303.424$ Bytes
Yottabyte	YB	$2^{80} = 1.208.925.819.614.629.174.706.176$ Bytes

tät von 4,7 GB in vielen Anwendungen korrekterweise nur die Kapazität 4,38 GB angezeigt.

$$10^9 = 1.000.000.000, \quad 2^{30} = 1.073.741.824$$

Der Kapazitätsunterschied zwischen Zweierpotenz und Zehnerpotenz ist in diesem Fall ca. 7,37 %.

Bei größeren Speichern ist der Unterschied noch größer. So bekommt ein Rechner mit einer angeblich 1 TB großen Festplatte meist nur ca. 930 GB angezeigt.

$$10^{12} = 1.000.000.000.000, \quad 2^{40} = 1.099.511.627.776$$

Der Kapazitätsunterschied zwischen Zweierpotenz und Zehnerpotenz ist in diesem Fall schon ca. 9,95 % und mit jeder weiteren Maßeinheit (PB, EB, ZB, usw.) wächst der Kapazitätsunterschied zwischen Zweierpotenzen und Zehnerpotenzen weiter.

Die International Electrotechnical Commission (IEC) schlug 1996 vor, die populären Größenfaktoren, die auf den Zweierpotenzen basieren, mit einem kleinen „i“ zu kennzeichnen und die etablierten Bezeichnungen der Maßeinheiten für die Zehnerpotenzen zu reservieren. Dieser Vorschlag konnte sich bislang nicht durchsetzen und die daraus resultierenden alternativen Bezeichnungen Kibibyte (KiB), Mebibyte (MiB), Gibibyte (GiB), Tebibyte (TiB), Pebibyte (PiB), Exbibyte (EiB) und Zebibyte (ZiB) sind außerhalb des akademischen Bereichs nicht besonders populär.

2.4 Informationsdarstellung

Daten sind Folgen von Nullen und Einsen, die beliebige Informationen repräsentieren. Um Texte und Zahlen durch Daten darzustellen, kodiert man die Zeichen des Alphabets (Groß- und Kleinschreibung), Satzzeichen wie Punkt, Komma und Semikolon, sowie einige Spezialzeichen wie zum Beispiel +, %, & und \$ in Bitfolgen. Zudem sind Steuerzeichen wie Leerzeichen (SP), Wagenrücklauf (CR) und Tabulator (TAB) nötig. Die am besten etablierte Kodierung ist der *American Standard Code for Information Interchange (ASCII)*.

2.4.1 ASCII-Kodierung

Die ASCII-Kodierung, häufig auch US-ASCII genannt, ist eine 7-Bit-Zeichenkodierung. Das heißt, dass jedem Zeichen eine Bitfolge aus 7 Bits zugeordnet ist. Es existieren also $2^7 = 128$ verschiedene Bitfolgen und exakt so viele Zeichen definiert die Zeichenkodierung (siehe Tab. 2.5). Von den 128 Zeichen sind 95 druckbare Zeichen und 33 nicht druckbare Zeichen. Die druckbaren Zeichen sind (beginnend mit dem Leerzeichen):

```
!"#$%&'() *+, - . / 0123456789 : ; < = > ?
@ABCDEFGHIJKLMNO PQRSTUVWXYZ [\ ] ^ _
`abcdefghijklmnopqrstuvwxyz { | } ~
```

Die nicht druckbaren Zeichen 00_{16} bis 20_{16} und $7F_{16}$, zum Beispiel Backspace (BS) und Carriage Return (CR) sind Steuerzeichen, die ursprünglich zur Steuerung eines Fernschreibers verwendet wurden. ASCII ist also nicht nur ein Standard zur Datenablage, sondern auch zur Datenübertragung geeignet. Den Anfang und das Ende einer Datenübertragung markiert man mit Start of Text (STX) bzw. End of Text (ETX). Die Steuerung der Übertragung ist mit nicht druckbaren Zeichen wie Acknowledge (ACK) und negative Acknowledge (NAK) möglich. Mit Bell (BEL) kann ein Sender, zum Beispiel bei einem Fehler, ein Alarmsignal an den Empfänger senden.

Das 8. Bit kann als Paritätsbit zur Fehlererkennung verwendet werden. In diesem Fall hat es den Wert Null, wenn die Anzahl der Einsen an den übrigen sieben Bitpositionen gerade ist und ansonsten den Wert Eins.

Durch verbesserte Protokolle zur Datenübertragung benötigt man das 8. Bit bei der Datenübertragung von ASCII-kodierten Texten nicht mehr zur Fehlererkennung. Darum wurde, um weitere Zeichen kodieren zu können, US-ASCII mit zahlreichen Erweiterungen zu einer 8-Bit-Zeichenkodierung erweitert. Wird jedem Zeichen eine Bitfolge aus 8 Bits zugeordnet, sind $2^8 = 256$ verschiedene Bitfolgen verfügbar. Es sind also 128 Zeichen mehr verfügbar, als das bei US-ASCII der Fall ist. Da diese 128 zusätzlichen Zeichen nicht ausreichen, um alle inter-