

8 Das Internet

Nachdem in den vorigen Abschnitten vor allem technische Aspekte von Rechnernetzen behandelt wurden, wollen wir uns jetzt dem wichtigsten globalen Netz widmen, dem *Internet*. Dieses Netz verbindet weltweit über 500 000 000 Rechner und eine um ein Vielfaches größere Zahl von Menschen und ist zu einer nie da gewesenen und alle Grenzen überwindenden Kommunikations-, Wissenschafts- und Wirtschaftsplattform geworden.

Historisch liegen die Anfänge des Internet in dem amerikanischen ARPANET (*Advanced Research Projects Agency-Net*) das seit 1969 als militärisches Netzwerk entwickelt wurde. Ab etwa 1972 wurde es auch eingesetzt, um Universitäten und Forschungseinrichtungen zu verbinden, die mit dem Verteidigungsministerium zusammenarbeiteten. Aus Sicherheitsgründen wurde das ARPANET später in einen öffentlichen und einen nichtöffentlichen Teil getrennt. Der öffentliche Teil wurde zum Internet.

Seit etwa 1990 hat sich das Internet durch die Einführung des World-Wide-Web (*WWW*) zu einem einfach zu bedienenden Informationssystem entwickelt. 1995 wurde es auch für kommerzielle Anwendungen geöffnet. Die Anzahl der Benutzer ist in den letzten Jahren drastisch gestiegen. Einem Abriss der Geschichte des Internets von R. H. Zakon (www.zakon.org) entnehmen wir die folgende Tabelle, die die gerundete Anzahl der mit dem Internet verbundenen *Hostcomputer* (engl. *host* = Gastgeber) im Verlauf der letzten 40 Jahre angibt. Neuere Daten sind erhältlich von www.isc.org. Die Zahlen aus jüngerer Zeit sind Schätzwerte, da viele Rechner immer nur temporär über Modem und Betreiber (*Provider*) am Internet angeschlossen sind und dabei nur eine zeitweilig genutzte *Hostadresse* ausgeliehen bekommen.

Zeitpunkt	Hosts	Zeitpunkt	Hosts	Zeitpunkt	Hosts
12/69	4	10/85	1.961	1/2000	72.398.000
10/70	11	11/86	5.089	1/2002	147.345.000
4/71	23	12/87	28.174	1/2004	233.101.000
1/73	35	10/88	56.000	1/2005	317.646.000
6/74	62	10/90	313.000	1/2006	394.992.000
3/77	111	10/92	1.136.000	1/2007	433.193.000
12/79	188	10/94	3.864.000	1/2008	541.677.000
8/81	213	7/96	16.729.000	1/2009	625.226.000
8/83	562	7/98	36.739.000	1/2010	721.243.000

Besser noch kann man sich die Entwicklung der Zahl der registrierten Hosts anhand der folgenden Grafik veranschaulichen. Sie hat einen logarithmischen Maßstab. Die Kurve ähnelt in einigen Teilen einer geraden Linie, ein Indiz für das exponentielle Wachstum des Internets in dieser Zeit. Etwa ab 1990 wurde das Internet von der Allgemeinheit entdeckt und hatte einen kräftigen Anstieg der Zahl der Hostcomputer zur Folge. In der Zeit um 2002 ist eine geringere Steigerung zu beobachten, eine Folge der Krise einer Reihe von Internetfirmen. Kurz danach erreichte die Steigerungsrate wieder ihre alten Werte. Seit etwa 2005 ist eine Stagnation auf hohem Niveau zu beobachten.

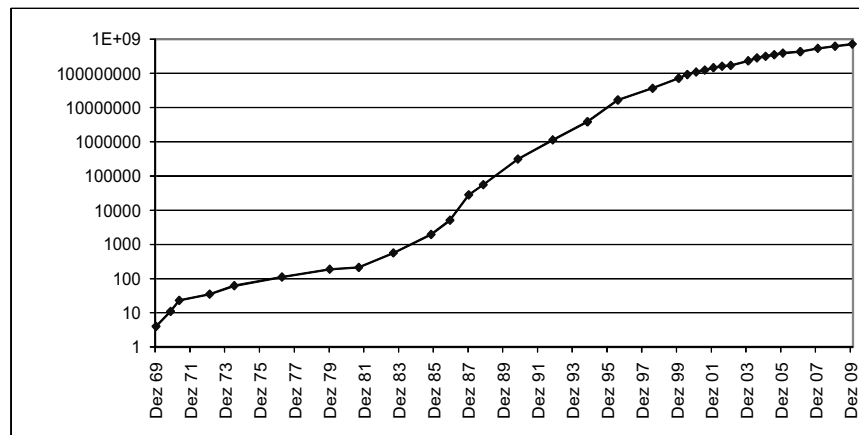


Abb. 8.1: Die Zahl der im Internet registrierten Hostcomputer

Die Zahl der Menschen, die das Internet benutzen, lässt sich ebenfalls nur schwer abschätzen, da einzelne Benutzer nicht als solche registriert sind. Die meisten haben zwar registrierte E-Mail-Adressen – über deren Zahl liegt allerdings keine Statistik vor.

8.0.1 Bildung von Standards im Internet

Bereits die ersten Schriftstücke aus dem Jahr 1969 mit Vorschlägen für das zukünftige, damals noch nicht so genannte, Internet wurden mit dem Kürzel *RFC* bezeichnet. Es steht für *request for comment* und ist eine Aufforderung an andere Entwickler, Kommentare zu diesem Arbeitspapier abzugeben. Praktisch alle Entwicklungen im Zusammenhang mit dem Internet sind in Form solcher RFCs dokumentiert. Diese sind durchnummeriert und über www.rfc-editor.org abrufbar.

Die wichtigste Funktion von RFCs ist die schrittweise Festlegung von Protokollen. Aus ersten Diskussionsentwürfen werden Vorschläge und schließlich Internet-Standards. So sind die aktuellen Versionen der Basisprotokolle IP und TCP durch die RFCs 791 bzw. 793 festgelegt. Die ursprünglichen Standards für E-Mail (SMTP) waren in den RFCs 821 und 822 festgehalten worden; sie wurden kürzlich überarbeitet und sind derzeit unter den Nummern 2821 und 2822 zu finden – mit dem Hinweis, dass diese Dokumente nunmehr die Vorgänger ersetzen.

Gelegentlich, bevorzugt mit Datum 1. April, finden sich nicht ganz so ernst gemeinte RFCs, wie z.B. RFC 1149 mit dem Titel „A standard for the transmission of IP datagrams on avian carriers“ (frei übersetzt: „Transport von IP-Paketen mit Brieftauben“).

Protokollnummern, Port-Nummern, Internetadressen und ähnliche „Zahlen“ wurden in der Frühzeit des Internet ebenfalls in RFCs geregelt. Da die Vergabe dieser „Zahlen“ jedoch ein dynamischer Prozess ist, wurde bereits im Dezember 1988 die IANA (*Internet Assigned Numbers Authority*) gegründet. Deren Direktor war der legendäre Internetpionier Jon Postel, der bis zu seinem Tod im Jahr 1998 auch als RFC Editor tätig war.

Heute ist IANA eine Unterabteilung der ICANN (*The Internet Corporation for Assigned Names and Numbers*), die dem Namen nach eine sehr ähnliche Aufgabe hat. IANA ist offenbar weiterhin für die Vergabe von Protokollnummern, Port-Nummern, Internetadressen und ähnliche *Zahlen* zuständig. ICANN beschäftigt sich darüber hinaus mit der Vergabe der im Internet verwendeten *Namen*. Diesen *Domain-Names* ist das Unterkapitel 8.3 gewidmet.

Für die Zuteilung von Namen und Adressen werden von IANA bzw. von ICANN so genannte *Registrare* eingesetzt. Auf internationaler Ebene ist das *INTERNIC* (www.internic.org) für die Vergabe von Domain-Namen und Internetadressen, die so genannte *Registrierung*, autorisiert. Daneben gibt es derzeit vier *regionale Internet Registrare* (RIR). Für Europa wurde eine Organisation namens *RIPE* (Réseaux IP Européens) eingesetzt. RIPE (www.ripe.net) ist derzeit außer für Europa noch für den Nahen Osten und für Teile von Afrika zuständig. Die anderen regionalen Registrare sind derzeit ARIN (*American Registry for Internet Numbers*, www.arin.net), APNIC (*Asia Pacific Network Information Centre*, www.apnic.net) und LACNIC (*Latin American and Caribbean IP address regional registry*, lacnic.net). Die regionalen Registrare erhalten Adressblöcke und Namensbereiche von IANA/ICANN. Hieraus ergibt sich eine hierarchische Vergabe von Internetadressen und Namen. Die unterste Ebene besteht aus den lokalen Registraren (LIR: *Local Internet Registry*), z.B. dem DENIC in Deutschland, dem AFNIC in Frankreich usw.

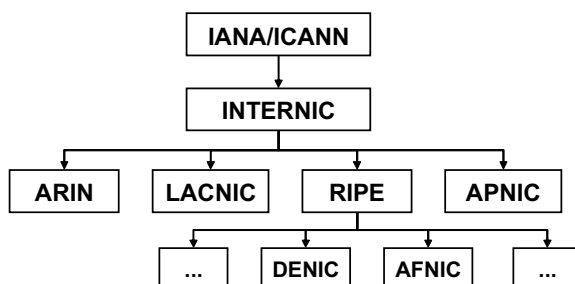


Abb. 8.2: Hierarchie der Registrare

8.1 Die TCP/IP Protokolle

Das Internet ist ein Netz von Netzen. Diese sind untereinander durch Vermittlungsrechner (*gateways* oder *router*) verknüpft. Basis des Internets ist eine Familie von Protokollen, die bis etwa 1982 spezifiziert wurden und unter dem Namen *TCP/IP* bekannt sind.

TCP (*transmission control protocol*) und *IP* (*internet protocol*) definieren zusammen das Übertragungsprotokoll des Internets. Sie sind auf Stufe 4 bzw. 3 des OSI-Referenzmodells angesiedelt. TCP ist verantwortlich für das Verpacken der zu übertragenden Daten in eine Folge von Datenpaketen. Diese werden mit einer Adresse versehen und an die niedrigere Schicht, IP, weitergereicht. IP ist für den Versand der einzelnen Pakete zuständig. Jedes Datenpaket wandert zu dem nächsten Vermittlungsrechner – meist wird dafür die englische Bezeichnung *Gateway oder Router benutzt* (siehe auch S. 617) – und wird von diesem weitergeleitet, bis es über viele Zwischenstationen an einem Vermittlungsrechner ankommt, in dessen Bereich (engl. *domain*) sich der Zielrechner befindet. Da die einzelnen Pakete einer Datei verschiedene Wege nehmen können, ist es möglich, dass sie in veränderter Reihenfolge beim Empfänger ankommen. Für die Zusammensetzung der Pakete in der richtigen Reihenfolge ist dann wieder TCP zuständig.

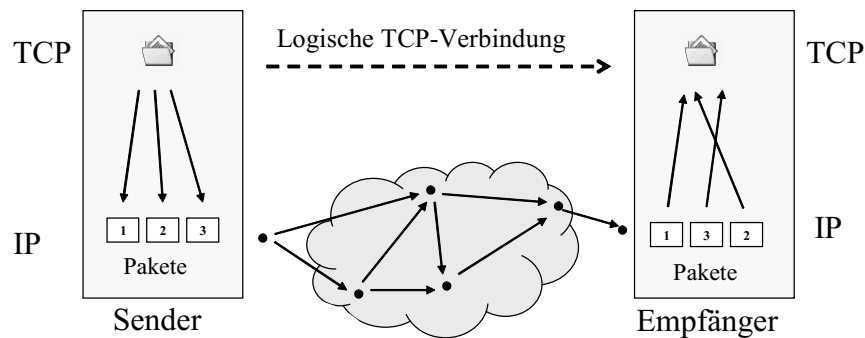


Abb. 8.3: TCP-Verbindung – realisiert durch IP

Ein wichtiges Ziel bei der Entwicklung von TCP/IP war es, ein ausfallsicheres Netz zu schaffen. Das ursprünglich militärische Netz sollte auch noch funktionieren, wenn durch einen Atomschlag einige Verbindungsrechner vernichtet sein sollten. Diese Ausfallsicherheit wird durch die Paketvermittlung gewährleistet. Allerdings steht dem auch ein großer Nachteil gegenüber. Die Pakete reisen unverschlüsselt über viele Vermittlungsrechner und können im Prinzip an jeder Zwischenstation gelesen werden. Zwar kann ein Benutzer selber seine Daten verschlüsseln, bevor er sie versendet, aber nicht immer ist ihm klar, dass er vertrauliche Daten auf die Reise schickt.

Daten werden von TCP/IP in Paketen verschickt. Die Daten stammen von Anwendungen bzw. von Protokollen, die bestimmte Anwendungen wie z.B. E-Mail unterstützen. In Abb. 8.4 sehen wir wie eine Datei mit dem FTP-Protokoll (siehe Abschnitt 8.5.3) verschickt wird. Die Datei wird dazu in kleine Einheiten gestückelt, da in einem Paket maximal 64 kB an Daten

versandt werden können. Meist beschränkt man sich jedoch auf wesentlich kleinere Pakete – z.B. 4 kB. Die gestückelten Daten werden jeweils in ein TCP-Paket eingepackt und mit Informationen versehen. Diese werden in einem *Header* zusammengefasst. Im nächsten Abschnitt werden wir uns näher damit befassen. Das TCP-Paket wird dann schließlich in ein IP-Paket eingepackt, um einen IP-Header erweitert und im Internet verschickt.

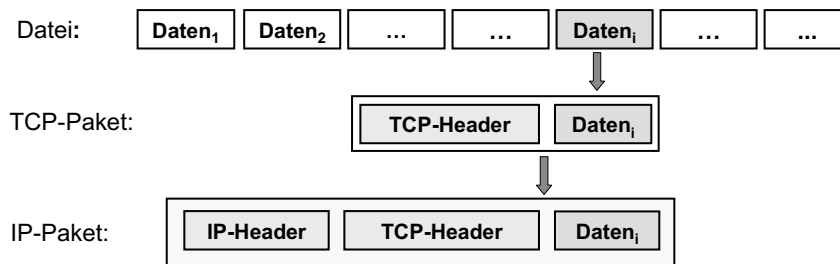


Abb. 8.4: Verpackung von Daten in Pakete

Die Abbildung könnte den Eindruck erwecken, als ob die Daten, also die Nutzlast, einen geringeren Anteil am Gesamtpaket hätten, als die beiden Header. Das ist natürlich i.A. nicht der Fall. Beide Header umfassen ca. 40 Byte. Bei einem Paket von mehr als 1 kB Gesamtgröße ist der Anteil des „Ballast“ also vergleichsweise gering.

8.1.1 Die Protokolle TCP und UDP

Typischerweise laufen auf einem Rechner viele Anwendungen gleichzeitig, die Daten über das Internet verschicken oder in Empfang nehmen. Während ich meine Post bearbeite, kann ich gleichzeitig surfen, Web-Radio hören und im Hintergrund eine große Datei herunterladen. Alle nötigen Daten kommen in vielen kleinen Paketen an bzw. werden als solche versandt. Wie bereits erwähnt, ist das IP-Protokoll für den Transport der Datenpakete vom Sender zum Empfänger verantwortlich. Der Versand erfolgt ungesichert und ohne Kenntnis, zu welcher Anwendung die Pakete gehören.

Die Verteilung der Pakete an die richtigen Anwendungen muss das TCP-Protokoll übernehmen. Dazu etabliert es so genannte *logische Verbindungen* zwischen spezifischen Anwendungen zweier Rechner, die längere Zeit andauern können. Jede erhält eine spezifische Nummer, die so genannte *Port-Nummer*. Der Begriff suggeriert einen Hafen (engl. *port*), in dem die zugehörigen Datenpakete einlaufen können.

Ports werden durch vorzeichenlose 16-Bit Zahlen, d.h. Werte im Bereich von 0 bis 65535 benannt. Von diesen sind einige für so genannte „*well-known ports*“ reserviert. Diese *gut bekannten Port-Nummern* werden benutzt, um die Dienste im Internet beliebigen Hosts unter einer definierten Adresse anbieten zu können. Für E-Mail (SMTP-Protokoll) wird beispielsweise die Port-Nummer 25 verwendet, für das HTTP Protokoll die Nummer 80. Ursprünglich waren 256 Nummern für diese Zwecke reserviert, mittlerweile sind es 1024.

Die Port-Nummern werden von der IANA vergeben bzw. registriert. Es werden derzeit drei Bereiche für Port-Nummern unterschieden. Der Bereich von 0 bis 1023 ist für die bereits genannten *well-known ports* reserviert. Der Bereich von 1024 bis 49151 wird von der IANA auf Wunsch für spezielle Anwendungen öffentlich registriert. Diese Registrierungen sind aber unverbindlich, d.h. diese Port-Nummern können auch anderweitig genutzt werden. Der Bereich von 49152 bis 65535 kann von Anwendungen ohne weitere Einschränkungen genutzt werden.

Jede TCP-Verbindung stellt eine zuverlässige Duplexverbindung zwischen zwei Anwendungen auf verschiedenen Hosts dar. In beiden Richtungen steht jeweils ein Übertragungskanal zum Lesen bzw. Schreiben zur Verfügung. Das TCP-Protokoll regelt den Verbindungsaufbau, den gesicherten Austausch von Daten und den Abbau der Verbindung im Falle eines normalen und eines fehlerbedingten Verbindungsendes.

Der Header eines TCP-Paketes ist eine Datenstruktur variabler Größe. Die Länge des Header ist in dem Feld HL (header length) angegeben. Die unterschiedliche Länge ergibt sich aus der Zahl der verwendeten Optionsfelder. Diese und einige weitere Felder in einem TCP-Header sind historisch bedingt und werden kaum verwendet.

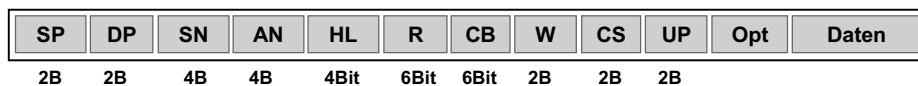


Abb. 8.5: Die Felder eines TCP-Paketes

Die Felder SP (source port) und DP (destination port) bezeichnen die Portnummern des sendenden und des empfangenden Programms. In dem Feld SN (sequence number) wird die Anzahl der Bytes angegeben, die in der Richtung des aktuellen Paketes gesendet worden sind und in AN (acknowledge number) wird die Anzahl der Bytes bestätigt, die in der umgekehrten Richtung erfolgreich angekommen sind. Diese Anzahlen verstehen sich kumulativ d.h. sie geben auch gleichzeitig eine Sequenznummer in dem Datenstrom in der jeweiligen Richtung an.

In dem W-Feld (Window) gibt der Sender an, wie groß die Differenz zwischen gesendeten und bestätigten Sequenznummern maximal werden darf. Man nennt diese Zahl auch *Fenstergröße* (engl. sliding window). Wenn diese zu groß ist, werden viele Daten auf Verdacht gesendet, bei einem Fehler müssen sie ggf. erneut gesendet werden. Wenn die Zahl zu klein ist, sinkt die Übertragungseffizienz, weil häufig auf eintreffende Bestätigungen gewartet werden muss. Die Übertragungsprotokolle berücksichtigen schließlich noch eine Prüfsumme im Feld CS (check sum) und einige der Bits aus dem Feld CB (Code-Bits).

Das TCP-Protokoll ist recht komplex. Es wurde konzipiert, um den gesicherten Transfer großer Datenmengen, z.B. für das FTP-Protokoll, zu garantieren oder um eine zeitlich länger andauernde Verbindung herzustellen, z.B. für das HTTP-Protokoll. In vielen Fällen reicht aber ein einfacheres Protokoll aus, wenn nur kleinere Datenmengen ungesichert übertragen werden sollen. Parallel zu TCP wurde daher *UDP (user datagram protocol)* entwickelt. Es wird häufig für einfache Anfragen benutzt, die ggf. wiederholt werden, wenn ein Fehler aufgetreten ist bzw. wenn nach einer bestimmten Zeit keine Antwort eingetroffen ist. Ebenso wie bei TCP wird bei UDP mit Hilfe von Port-Nummern eine direkte Verbindung zwischen zwei

Anwendungen hergestellt. Ein Beispiel für die Anwendung von UDP sind die später diskutierten Anfragen an Domain-Name-Server. Der Header eines UDP-Paketes besteht aus genau 4 Feldern, ist also einfacher als der eines TCP-Paketes. Außer den Feldern SN, DN und CS gibt es nur noch ein Feld, das die Gesamtlänge des Paketes angibt.

8.1.2 Das IP Protokoll

Das IP-Protokoll ist für den Versand *einzelner* Pakete verantwortlich. Jedes Paket wird *nach besten Kräften* (engl. *best effort delivery*) zu dem Empfänger befördert. Der wesentliche Inhalt des Headers eines IP-Paketes sind daher Absender- und Zieladresse. Mit diesen werden wir uns im nächsten Unterkapitel beschäftigen.

Derzeit wird überwiegend die Version 4 des IP-Protokolles benutzt, abgekürzt *IPv4*. Die Zahl der mit diesem Protokoll adressierbaren Rechner bzw. Netze hat sich als zu klein erwiesen. Daher wurde ein neues Protokoll, *IPv6* (IP Version 6) entwickelt. Die Versionen 4 und 6 des IP-Protokolles (über eine Version 5 ist wenig bekannt) unterscheiden sich hauptsächlich in der Adressbreite. Während IPv4 mit 32-Bit-Adressen arbeitet, sind bei der neueren Version 128 Bits für Adressen vorgesehen. Die Anzahl der mit IPv6 adressierbaren Rechner bzw. Netze ist offensichtlich sehr groß – Spötter sagen, man könnte damit jedes Atom unseres Planeten adressieren.

Die Einführung von IPv6 wäre im Prinzip schon ab 1995 möglich gewesen. Seither liest man in vielen Publikationen (so auch in den letzten Auflagen dieses Buches) immer wieder den Satz: „Die Umstellung von IPv4 auf IPv6 wird schrittweise in den nächsten Jahren erfolgen.“ Tatsächlich wird IPv6 aber immer noch nicht in nennenswertem Umfang eingesetzt. Auch gibt es keine Pläne, die eine endgültige Umstellung bzw. Einführung von IPv6 vorsehen. Zweifellos hätten 48- oder 64-Bit-Adressen ausgereicht, um die Schwächen der IPv4 Adressierung zu überwinden – vielleicht wären solche auch schneller allgemein akzeptiert worden.

IP-Pakete werden häufig *Datagramme* genannt. Ihre Länge muss ein Vielfaches von einem 4 Byte Wort sein. Dies gilt auch für den Header. Ähnlich wie bei TCP-Paketten können optionale Felder benutzt werden. Daher können die Header unterschiedlich lang sein – mindestens aber 5 und höchstens 16 Worte. Die Gesamtlänge ist maximal 65 536 Bytes = 16384 Worte.



Abb. 8.6: Struktur von IPv4 Paketen

Das erste Feld eines IPv4 Headers gibt Auskunft über die Version des Protokolls, enthält also eine 4. Die meisten Felder sind historisch bedingt und spielen derzeit kaum noch eine Rolle. Die wichtigsten Felder sind SA (Source Address), DA (Destination Address), TTL (Time to live) und PROT. Das letztere Feld gibt Auskunft über das enthaltene Protokoll. Im Falle eines eingebetteten TCP-Paketes enthält dieses Feld z.B. die Nummer 6. Die Felder SA und DA enthalten die IPv4-Adressen von Absender und Empfänger. Das Feld TTL enthält einen Zähler, der jedes Mal, wenn das Paket von einem Vermittlungsrechner weitergeleitet wird, um 1 erniedrigt wird. Wenn das Feld den Wert 0 erreicht, wird das Paket verworfen. Dadurch sollen