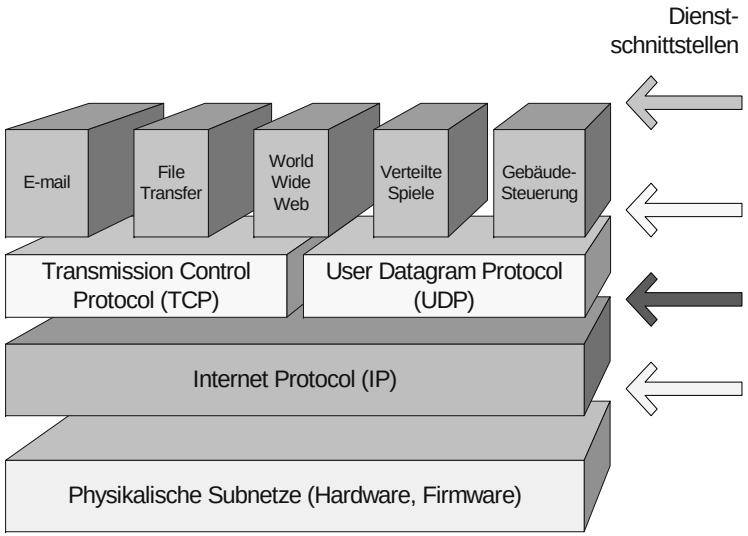
Einführung: Grundlegende Design-Prinzipien des Internet

Schichtenarchitektur von TCP/IP



Ursprüngliche Entwurfsentscheidungen

- Paketvermittlung
  - damals eine neue Technik, im ARPANET erforscht
- Hohe Verfügbarkeit des Netzes
  - Datengrammtechnik (vs. virtual circuit)
  - wenig Zustandsinformation im Netz
  - verteilte Verwaltung
- Soll viele verschiedene Dienste unterstützen können
  - Internet Protocol (IP) als Netzwerkprotokoll
  - Transmission Control Protocol (TCP), User Datagram Protocol (UDP) als Transportprotokolle (>Bezeichnung TCP/IP)
  - Anwendungen direkt auf Transportprotokolle aufgesetzt
- Netzverbund mit heterogenen Teilnetzen
  - Minimale Anforderungen an die Subnetze (bez. Zuverlässigkeit, Durchsatz, etc.), Fragmentierungsfunktion in IP
- „Offenes System"
  - Spezifikation offen gelegt und unter öffentlicher Kontrolle

Dienstschnittstellen

- zur Anwendung: nicht standardisiert, anwendungsabhängig
- zu den Transportprotokollen: De-facto Standard (socket-Schnittstelle)
  - Dienst von UDP: Verbindungslos, unzuverlässig
  - Dienst von TCP: Verbindungsorientiert, zuverlässig
- zu IP: Eingschränkt auf Super-User / Systemprogramme, via raw socket
  - Verbindungslos, unzuverlässig
- zu den phys. Subnetzen: Netz- und implementationsabhängig, oft jedoch IEEE 802.x

Geschichte des Internet (I)

- Baut auf Forschung im Bereich Paketvermittlung auf (Arpanet, ca. ab 1967)
- 1973: Bob Kahn postuliert eine neue Architektur, basierend auf Konzept Netzverbund
- 1973/74: Implementation des Konzepts in der Gruppe von Vint Cerf in Stanford; erstes Internet mit 3 Netzen
- 1977: Einbindung des Arpanet
- Ab ca. 1980: Arpanet ist wichtiger Backbone des wachsenden Internet
- 1983: TCP/IP als Standard für das US DoD verankert
- DARPA finanzierte Implementation von TCP/IP (entwickelt von BBN) und deren Integration in Berkeley UNIX (BSD); socket Schnittstelle.

Standardisierungsprozess: Dokumentation

- Request for Comment (RFC): Reihe von elektronisch zugänglichen Publikationen, welche das Internet beschreiben.
- Frühere Reihe von Publikationen: Internet Engineering Notes (IEN). Nicht mehr weitergeführt
- Internet Drafts (ID): Diskussions- und Entwurfsdokumente für die Standardisierung, zeitlich beschränkge Gültigkeit.
- Dokumentation ist vollständig im Internet verfügbar.

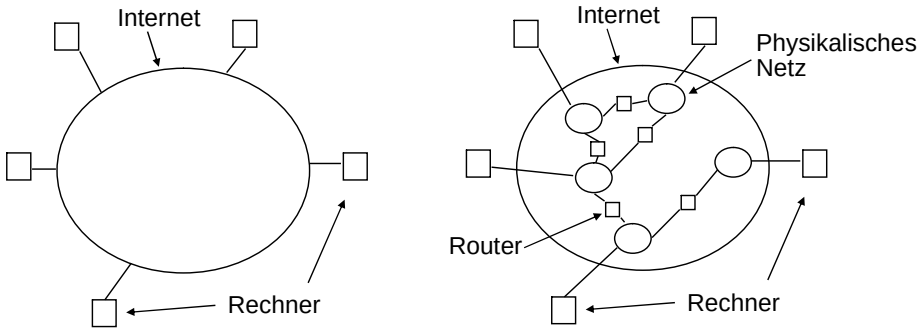
Geschichte des Internet II

- 1986: NSFNET in USA als Rückgratnetz von neuen regionalen Netzen
- Ab 1987: Aufbau von TCP/IP Netzen in Europa und Australien- weltweites Internet
- ca. 1988/89: Aufbau und Anschluss des Schweizerischen akad. Forschungsnetzes (SWITCH) an das Internet. Rückgratnetz mit 2 Mbit/s!
- 1990: Übergang von 56 kbit/s auf 1.544 Mbit/s (T1) für Leitungen des NSFNET
- 1990: Mehr als 3000 Netze und 200'000 Hosts
- 1992: Übergang auf 45 Mbit/s im Backbone im NSFNET; Start der Entw. von *IP next generation*
- 1994/95: Explosion des Internet nach der „Erfindung" des WWW
- Seither: Triebfeder der Informationsgesellschaft

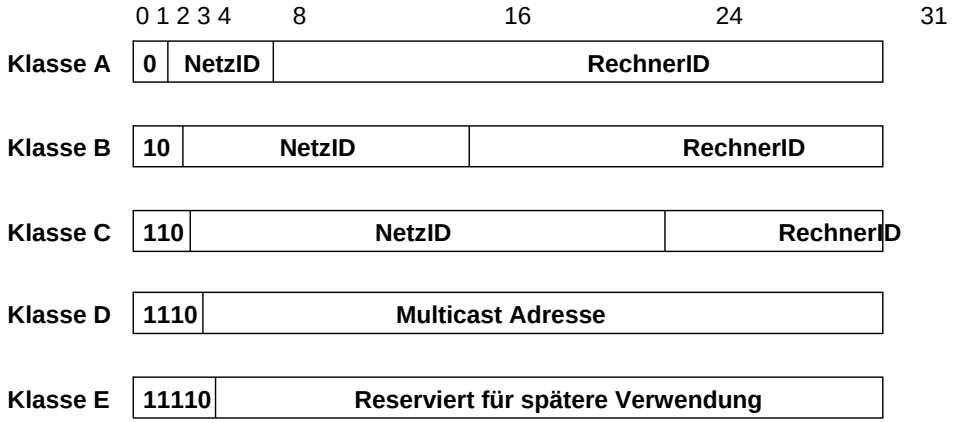
Adressierung
Address Resolution Protocol
Das IP Protokoll

Konzept und Architekturmodell des Internet

"The TCP/IP internet protocols treat all networks equally. A local area network like an Ethernet, a wide area network like the NSFNET backbone, or a point-to-point link between two machines each count as one network." (D. Comer)



Adressierung im Internet: Adresstypen



Konzept und Architekturmodell des Internet IV

- **Das Internet beruht auf dem Zusammenschluss von teilautonomen Subnetzen mittels Verbindungsrechnern (Routern).**
- **Router leiten den Datenverkehr gemäss einer Netzwerkadresse, nicht einer Endsystemadresse.**
- **Diese Form der Datenweiterleitung ist transparent für Benutzer.**

Adressierung im Internet: Darstellung

- Darstellung als 4 Oktette in Dezimalnotation, getrennt durch einen Punkt, z.B. 129.132.66.1

- Klasse A zwischen 1 und 126
- Klasse B zwischen 128.1 und 191.254
- Klasse C zwischen 192.1.1 und 223.254.254

- Um ein Netz zu adressieren, wird der Rechnerteil einer Adresse auf Null gesetzt, z.B. B-Netz der ETH: 129.132.0.0
- Um alle Rechner in einem Netz zu erreichen (Broadcast), wird der Rechnerteil auf 1 gesetzt, z.B. 129.132.255.255

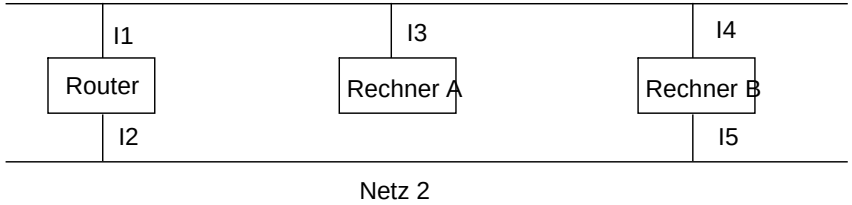
Adressierung im Internet: Spezielle Adressen

- **Eine Null in einem Teil der Adresse bezeichnet per Konvention den lokalen Rechner bzw. das lokale Netz.**
- **Die Adresse 127.0.0.0 (oft auch 127.0.0.1) ist der "lokale loopback" eines Rechners.**

Alles Null		**Lokaler Rechner**
Alles Null	Rechner	**Rechner auf lokalem Netz**
Alles Eins		**Beschränkter Broadcast (auf lokalem Netz)**
Netz	Alles Eins	**Gerichteter Broadcast für "Netz"**
127	Beliebig	**Lokaler Loopback**

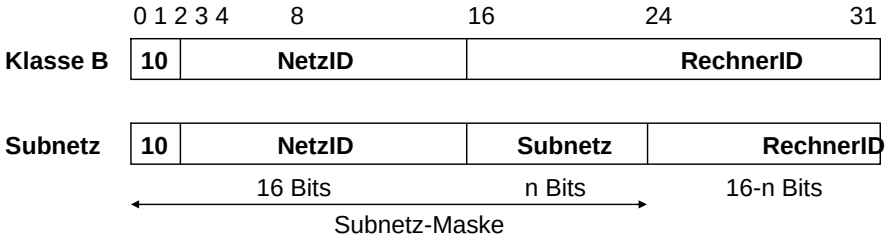
Adressierung im Internet: Schwachpunkte

- Wenn ein Rechner an ein anderes Netz angehängt wird, muss seine Adresse geändert werden.
- Die Reihenfolge der Adressbytes ist im Standard festgelegt.
- Wenn z.B. ein C-Netz auf mehr als 255 Rechner wächst, müssen alle Rechner auf ein B-Netz migriert werden.
- Ein Rechner mit mehreren Anschlüssen an das Internet braucht mehrere Adressen, die auch verschiedene Routen implizieren.



Adressierung im Internet: Namensautorität und Subnetze

- Die Adressautorität im Internet wird durch die zentrale Vergabe von Netzadressen durch das NIC in den USA ausgeübt.
- Um die Freiheit der lokalen Konfiguration zu erhöhen, und die Anzahl vergebener Netzadressen zu minimieren, ist die Verwendung lokaler Subnetz-Masken zur internen Unterteilung des Rechner-Teils der Adresse möglich.



Beispiel: Netz 129.132.0.0, Maske 255.255.255.192 = 10 Bit Subnetz

Problem der Adressabbildung

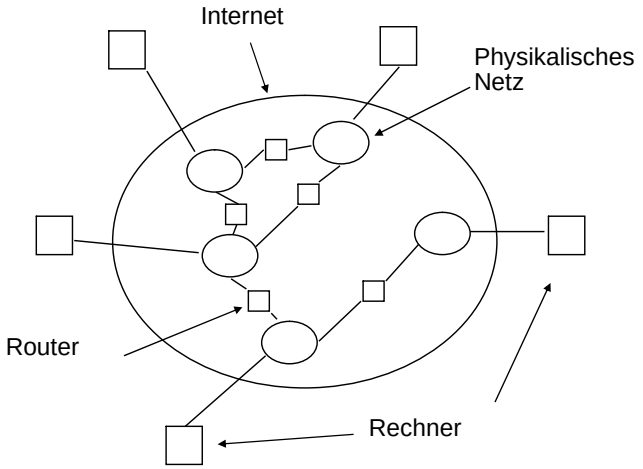
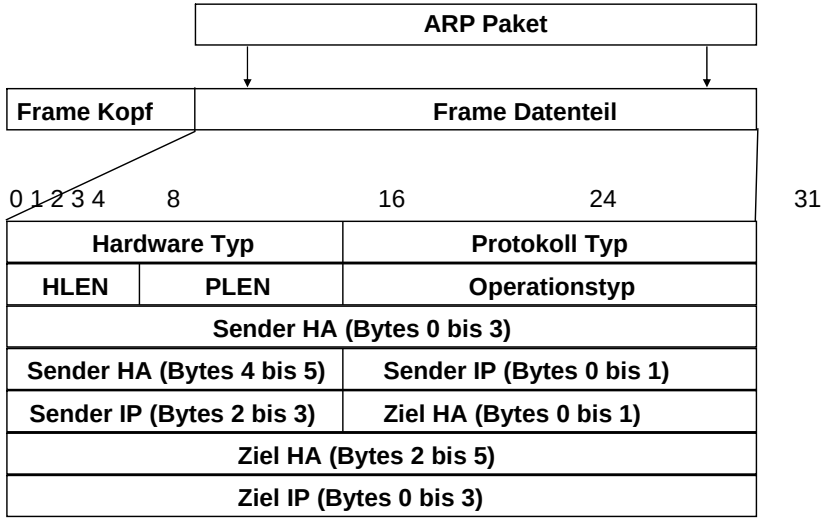


Abbildung von IP-Adressen auf physikalische Adressen

- Gegeben 2 Rechner, die am selben physikalischen Subnetz angeschlossen sind.
- Beide Rechner haben je eine IP-Adresse, und je eine physikalische Adresse bezüglich ihres gemeinsamen Netzes (z.B. eine Ethernet-Adresse).
- Will Rechner A Daten an Rechner B senden, so muss er anhand der IP-Adresse von Rechner B die Ethernet-Adresse von Rechner B herausfinden, um die Daten über das gemeinsame Netz zu senden.

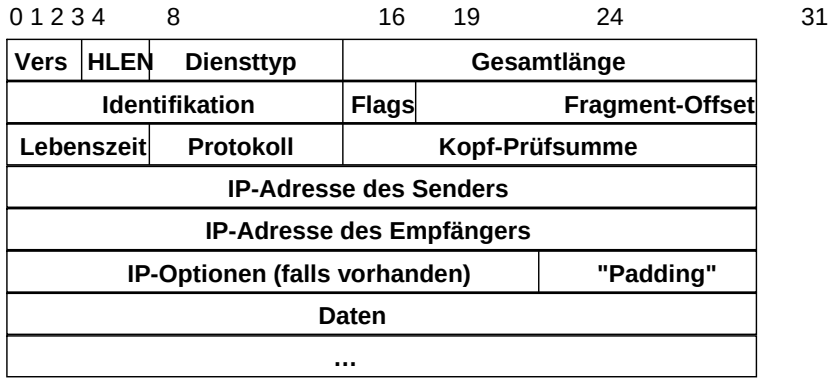
Aufbau eines ARP-Paketes



Wege zur Abbildung von IP-Adressen auf physikalische Adressen

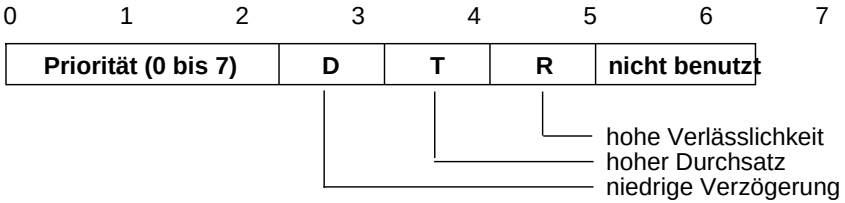
- Direkte Abbildung durch Berechnung aus der IP-Adresse: dies ist nur in bestimmten Netzen möglich, und auch nur, solange das Adressierschema in beiden Adressräumen eingehalten wird.
- Suche der physikalischen Adresse in einem Verzeichnisdienst anhand der IP-Adresse.
- Dynamische Bindung durch Nachfragen auf dem lokalen Netz mittels des "Address Resolution Protocol": Rechner A sendet ein spezielles Broadcast-Paket auf das lokale Netz, in dem die IP-Adresse von Rechner B angegeben ist, und in dem nach der physikalischen Adresse von Rechner B gefragt wird. Rechner B füllt die gesuchte Adresse ein und sendet das Paket zurück.

Aufbau eines IP-Paketes

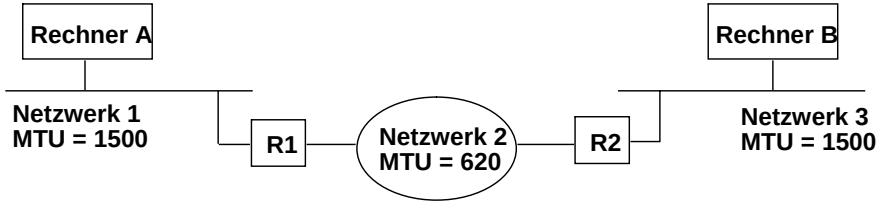


Länge und Diensttypangabe in einem IP-Paket

- Das Feld HLEN gibt die Länge des IP-Kopfes an. Das einzige variabel lange Feld sind die IP-Optionen, die mittels des "Padding-Feldes" wieder auf ein Vielfaches von 32 gebracht werden.
- Typische Grösse (ohne Optionen) ist 20 Byte (HLEN = 5).
- Diensttyp:



Fragmentierung von IP-Paketen



Datagram-Kopf	1400 Bytes Daten			
Datagram-Kopf	600 Bytes Daten	600 Bytes Daten	200 Bytes Daten	
Fragment-1-Kopf	600 Bytes Daten	**Flag "weitere Fragmente" gesetzt**		
Fragment-2-Kopf	600 Bytes Daten	**Flag "weitere Fragmente" gesetzt**		
Fragment-3-Kopf	200 Bytes Daten			

Weitere Felder in einem IP-Paket: Fragmentierung

- Die Felder Identifikation, Flags und Fragment-Offset kontrollieren die Zerlegung zu langer Datenpakete in mehrere kleinere Pakete.
- Ethernet-Frames können maximal 1500 Bytes lang sein. Diese Maximalwerte werden als MTU (maximum transfer unit) bezeichnet.
- Ist ein IP-Paket grösser als die vorhandene MTU, muss das IP- Paket fragmentiert (zerlegt) werden.

Weitere Felder in einem IP-Paket: Identifikation, Offset, Flags

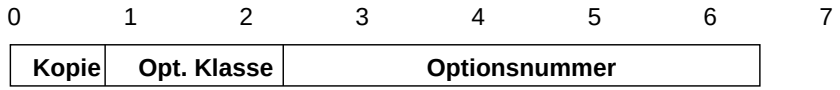
- Das Feld "Identifikation" enthält eine eindeutige Nummer des ursprünglichen IP-Pakets.
- Der Fragment-Offset spezifiziert die Stelle im ursprünglichen IP- Paket, an dem das aktuelle Fragment eingesetzt werden muss.
- Mittels des Felds "Flags" kann Fragmentierung verboten werden, es wird auch zum Signalisieren weiterer Fragmente benutzt.
- Das Feld "Gesamtlänge" in einem Fragment bezieht sich auf die Länge des Fragmentes, nicht auf die Länge des IP-Pakets.
- Einmal fragmentierte Pakete werden erst beim Empfänger wieder zusammengesetzt (Nachteile: Zusatzlast und Gefahr von Verlust).

Weitere Felder in einem IP-Paket: Lebenszeit

- Das Feld "Lebenszeit" gibt an, wie lange (in Sekunden) ein Paket im Internet unterwegs sein darf, bevor es gelöscht wird. Beim Erstellen des Paketes wird eine Maximalzeit angegeben, die bei jeder Weiterleitung des Pakets dekrementiert wird.
- Wird ein Paket in einem Router verzögert, wird ein entsprechend höherer Wert abgezogen.
- Wird ein Paket wegen "Lebenszeit = 0" vor seiner Ankunft beim Empfänger gelöscht, muss das löschende System eine Fehler- meldung an den Urheber des Pakets zurücksenden.

Weitere Felder in einem IP-Paket: IP-Optionen

- IP-Optionen



- Das Flag "Kopie" gibt an, ob bei Fragmentierung die Optionen nur im ersten Fragment, oder in allen Fragmenten gesetzt werden.
- Optionsklasse

0	Normales Datengramm oder Netzwerk Kontrolle
1	reserviert für zukünftige Benutzung
2	Fehlersuche und Messungen
3	reserviert für zukünftige Benutzung

Weitere Felder in einem IP-Paket: Protokoll, Prüfsumme, Adressen

- Das Feld "Protokoll" gibt an, welches hierarchisch über IP liegende Protokoll das Paket erzeugt hat, d.h. in welchem Format sich die Daten befinden.
- Das Feld Kopf-Prüfsumme dient der Datensicherung, bei der Bildung der Prüfsumme wird dieses Feld als "0" angenommen.
- Die Felder mit den IP-Adressen von Sender und Emfänger haben End-zu-End-Signifikanz, d.h sie werden nicht verändert, während das Paket durch das Internet transportiert wird.

Weitere Felder in einem IP-Paket: Optionale Elemente

- Optionen für Leitweglenkung und Zeitstempel
- Die Option "Wegaufzeichnung" ermöglicht die Protokollierung des Weges des Pakets durch das Netz im Optionsfeld des IP-Pakets.
- Die Option für die Wahl des Leitweges ermöglicht es dem Sender eines IP-Pakets, den Weg zum Empfänger zu diktieren. Diese Option wird meist zu Testzwecken benötigt.
- Die Zeitstempel-Option arbeitet ähnlich zur Option "Wegaufzeich- nung", es wird jedoch ein zusätzlicher Zeitstempel angegeben. Weitere Flags steuern die Details der Angabe von Zeitstempeln.

Internet Control Message Protocol (ICMP)

ICMP-Meldungstypen

- 0 Echo-Antwort
- 3 Destination unerreichbar
- 4 "Source Quench"
- 5 Änderung einer Route
- 8 Echo-Anforderung
- 11 Datagram-Lebenszeit überschritten
- 12 Parameter-Problem im Datengramm
- 13 Zeitstempel-Anforderung
- 14 Zeitstempel-Antwort
- 15 Informations-Anforderung (nicht mehr benötigt)
- 16 Informations-Antwort (nicht mehr benötigt)
- 17 Adressmasken-Anforderung
- 18 Adressmasken-Antwort

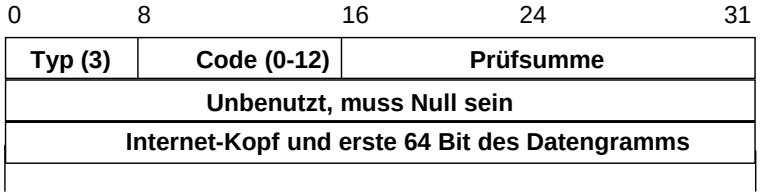
Einführung in ICMP

- Zur Steuerung und Verwaltung des Internet ist ein gesondertes Protokoll nötig, welches "normalen" Benutzern nicht zugänglich ist.
- Typische Aufgaben des Internet-Kontrollprotokolls sind:
  - Koordination zwischen Routern und Endsystemen
  - Fehlererkennung und -korrektur
  - Überwachung und Messung des Verkehrsaufkommens
- ICMP stellt eine Kommunikationsmöglichkeit zwischen der IP-Software auf Internet-Rechnern zur Verfügung.
- ICMP benutzt den IP-Dienst, gehört logisch aber auf die selbe Protokollschicht, wie das IP-Protokoll.

Meldungen über nicht erreichbare Destinationen

- Wenn ein IP-Paket nicht weitergeleitet werden kann, wird eine entsprechende Fehlermeldung erzeugt.

0	Netz nicht erreichbar	7	Zielrechner unbekannt
1	Rechner nicht erreichbar	8	Zielrechner isoliert
2	Protokoll nicht erreichbar	9	Netzkommunikation unerwünscht
3	Port nicht erreichbar unerwünscht	10	Rechnerkommunikation
4	Fragmentierung benötigt unerreichbar	11	Netz für diesen Dienst
5	Falsche Quell-Route unerreichbar	12	Rechner für diesen Dienst
6	Zielnetz unbekannt		



Automatische Konfiguration mit dem Dynamic Host Configuration Protocol (DHCP)

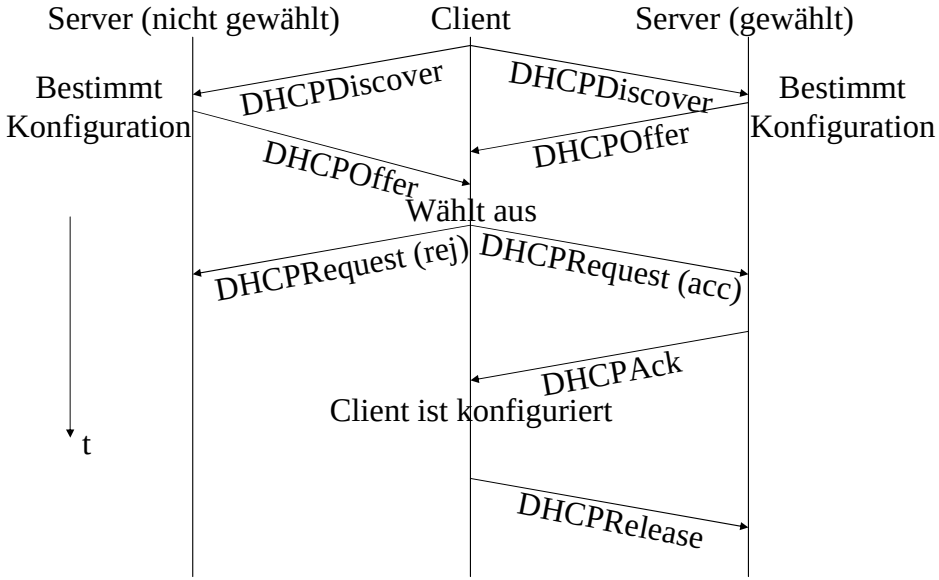
Aufbau einer DHCP-Nachricht

op-Operation	htype	hlen	hops
xid-Transaction ID			
secs-Sekunden seit Beginn		Flags	
ciaddr-Client-Adresse (falls schon zugewiesen)			
yiaddr-Neu zugewiesene Client-Adresse			
siaddr-Adresse des Bootstrap-Server			
giaddr-Adresse des Relay-Agent			
chaddr-Hardware-Adresse des Client (16)			
sname-Hostname des Servers (optional) (64)			
file-Bootstrap-Filename (128)			
Optionen (variabel lang)			

Wozu DHCP?

- Automatische Konfiguration von Hosts
- Zu konfigurierende Parameter:
  - IP-Adresse
  - Gateway- (Router-) Adressen
  - IP-Maske (Subnetting)
  - Adressen der DNS-Server
  - Link MTU, default time-to-live
  - … und sehr viel mehr, s. Appendix A von RFC 2131
- Ohne DHCP: Einstellung von Hand (Netzwerk-Kontrollfeld in Windows, Registry, oder nicht einstellbar
- Unangenehm: Bei einer Konfigurationsänderung von Hand muss Windows (95, 98) neu gestartet werden

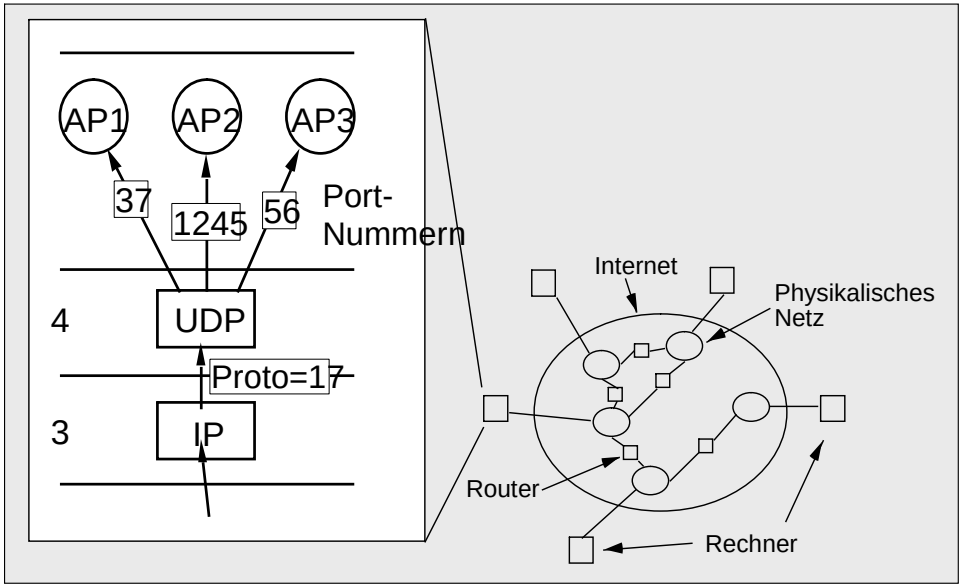
Typischer Ablauf des DHCP-Protokolls



Die Transportprotokolle:

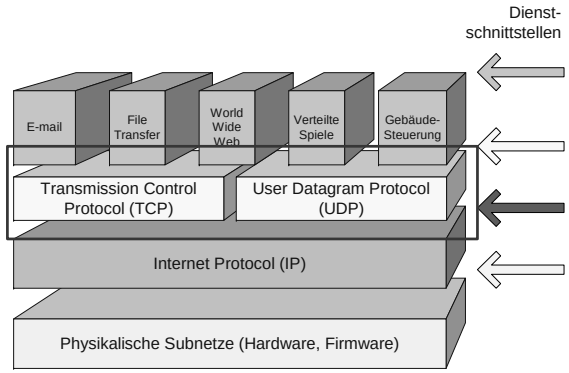
Transmission Control Protocol (TCP) User Datagram Protocol (UDP) Die Socket-Schnittstelle

Adressierung von Anwendungsprozessen: Beispiel TCP/IP - Portnummern



Einführung in TCP

- **TCP implementiert ein verbindungsorientiertes, zuverlässiges Transport- Protokoll, aufbauend auf dem IP-Dienst.**



Koordination der global zugeordneten Ports

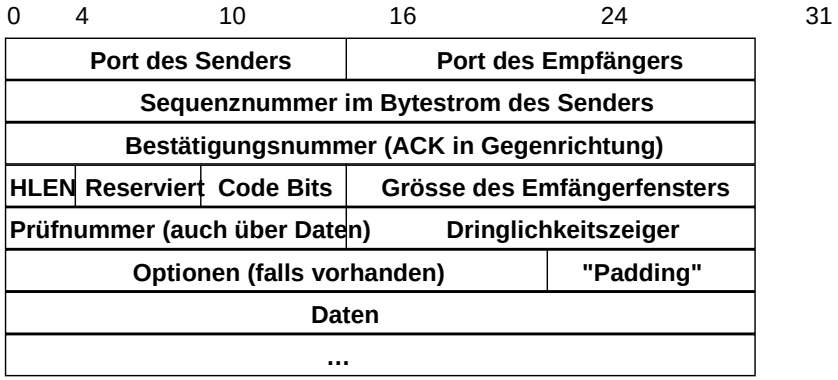
- Internet Assigned Numbers Authority (IANA): Zuständig für Vergabe von Konstanten in TCP/ IP-Protokollen (port numbers, protocol numbers, ...) – neu unter der Verantwortung von ICANN
- Bereich 0.. 1023: Für globale "well known" ports, kontrolliert von der IANA
- Bereich 1024 .. 65535: Frei für dynamische Allozierung durch Prozesse oder für statische Allozierung mit lokaler Bedeutung
  - Registrierung durch IANA ist optional
- Aktuelle globale / statische Zuordnungen: http://www.iana.org/assignments/port-numbers

Well-known port numbers: /etc/services (Auszug)

```
# Note that it is presently the policy of IANA to assign a single well-known
# port number for both TCP and UDP; hence, most entries here have two entries
# even if the protocol doesn't support UDP operations.
# Updated from RFC 1700, “Assigned Numbers”

echo            7/tcp
echo            7/udp
discard         9/tcp           sink null
discard         9/udp           sink null
ftp-data        20/tcp
ftp             21/tcp
telnet          23/tcp
smtp            25/tcp          mail
time            37/tcp          timserver
time            37/udp          timserver
nameserver      42/tcp          name            # IEN 116
whois           43/tcp          nicname
domain          53/tcp          nameserver      # name-domain server
domain          53/udp          nameserver
```

TCP-Segmentformat



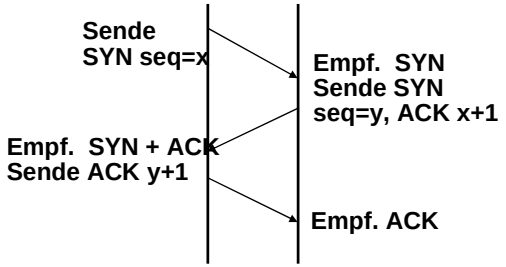
- Code Bits: URG, ACK, PSH, RST, SYN, FIN
- Dringlichkeitszeiger: zeigt auf das Ende der dringenden Daten im TCP-Datenfeld.

Eigenschaften des Transmission Control Protocol (TCP)

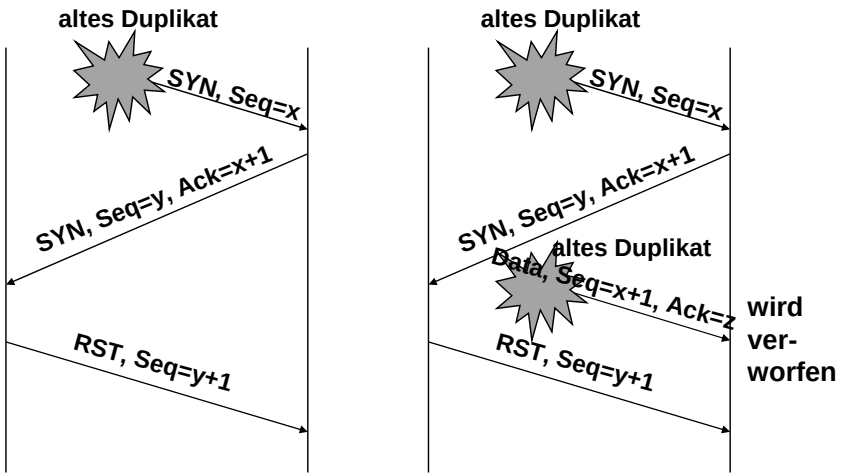
- verbindungsorientiert
- Vollduplex-Verbindung
- stellt eine “byte pipe” zur Verfügung - unstrukturierter Datenstrom
- Sliding Window-Protokoll
- Folgenummern sind Bytenummern
- Maximale Fenstergrösse 2^{16} Bytes
- Variable Grösse des Sendefensters bestimmt durch das Maximum von:
  - Angabe des Empfängers (receiver window size)
  - Congestion window size, abhängig von einer lokalen Schätzung der Netzbelastung -> “Slow Start” Algorithmus

Verbindungsaufbau

- Aktives Öffnen einer Verbindung (SYN)
- Passive Seite nimmt eine Verbindung auf einer bestimmten Port-Nummer entgegen
- Die initialen Sequenznummern werden auf jeder Seite zufällig gewählt und bestätigt.
- 3-fach-Handshake (nötig wegen des unzuverlässigen Dienstes von IP):

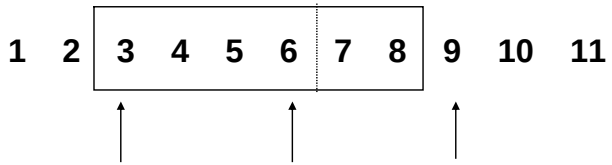


Verbindungsaufbau, zwei Fehlerszenarien



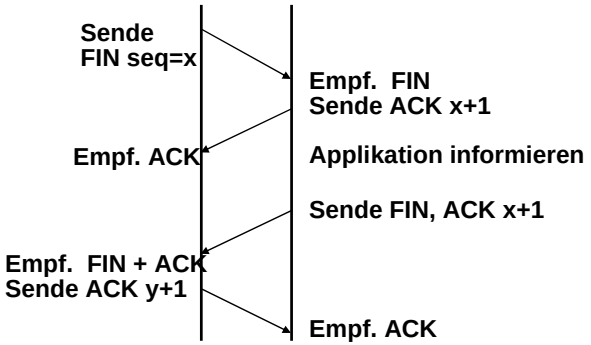
Segmente, Datenströme und Sequenznummern

- TCP betrachtet einen Datenstrom als Sequenz von Bytes, die für die Übertragung in TCP Segmente eingeteilt werden. Jedes Segment wird dann in der Regel auf ein IP Paket abgebildet. (Grösse eines Segmentes bei lokaler Übertragung gemäss physikalischem Netz, sonst 536 Bytes)
- TCP verwendet ein "sliding window" Protokoll, um möglichst effizient Daten zu übertragen, und Flusskontrolle zu ermöglichen. Bei einer Vollduplex Verbindung müssen insgesamt 4 Fenster verwaltet werden.



Abbau einer TCP-Verbindung

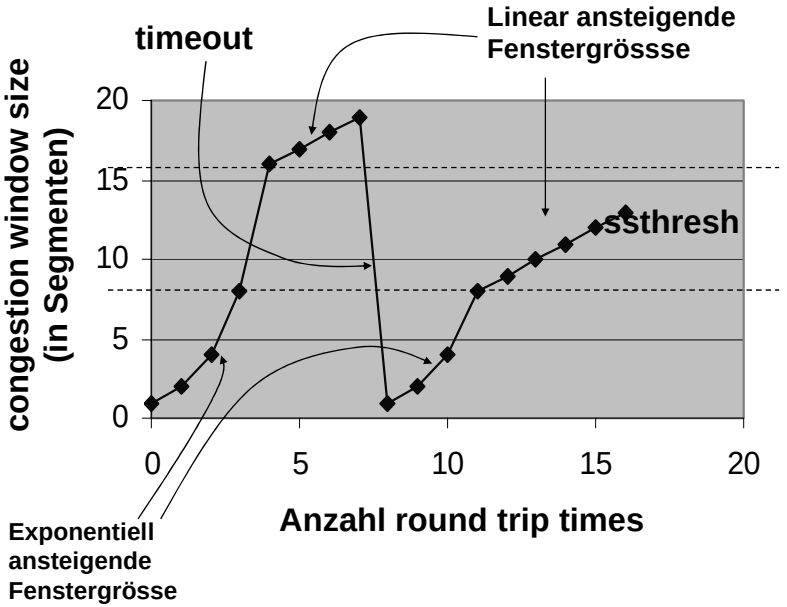
- Aktive Seite (links) schliesst Verbindung mit FIN-Flag
- Neue Daten werden nicht mehr übertragen, von rechts ankommende Daten werden jedoch noch bestätigt.
- 4-fach-Handshake; jede Seite wird separat beendet (TCP half close)



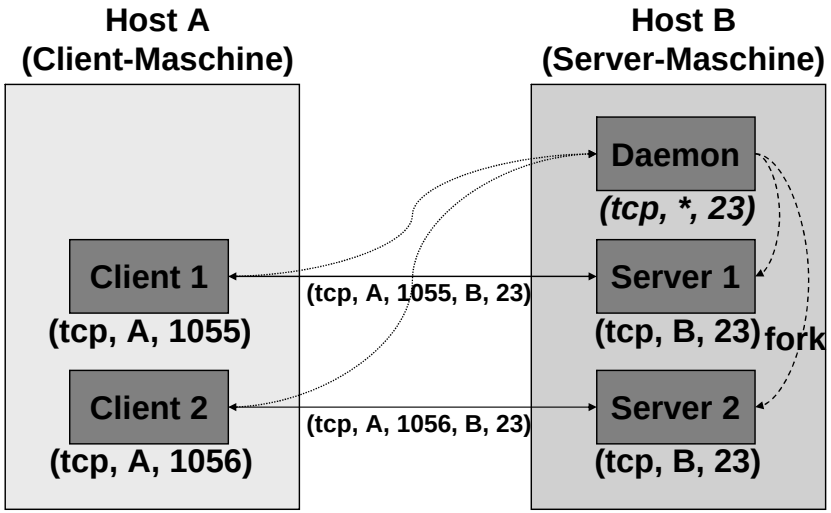
Variable Fenstergrössen

- Die Fenstergrösse im TCP "sliding window" Protokoll kann variiert, d.h. an den "Füllstand" des Netzes bzw. des Empfängers angepasst, werden.
- Flusssteuerung
  - Jedes Bestätigungspaket enthält einen "window advertisement" Wert, in dem der Empfänger angibt, für wieviele weitere Pakete er noch freie Kapazität hat (das Fenster kann also grösser oder kleiner werden).
- Verkehrssteuerung
  - Jacobsen's "slow start" Algorithmus variiert die Grösse des Sendefensters, um die Senderate an die Netzbelastung anzupassen (s. Folie 18).

Slow Start Algorithmus



Identifikation von Verbindungen



Verbindungen und Verbindungsendpunkte

Eine TCP-Verbindung wird durch ein Paar von Adressen und Port-Nummern identifiziert (Verbindungsendpunkte):

- IP-Adresse und Port-Nummer Host A
- IP-Adresse und Port-Nummer Host B

Jede Verbindung wird durch ein Paar von Verbindungsendpunkten eindeutig identifiziert -> mehrere Verbindung zwischen den gleichen Hosts sind dadurch gleichzeitig möglich.