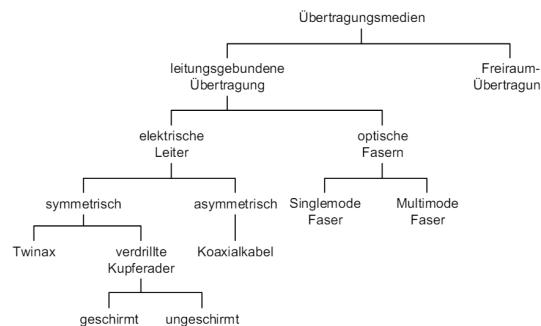


Übertragungsmedien

Kategorisierung Übertragungsmedien



Signale

Ausbreitungsgeschwindigkeit

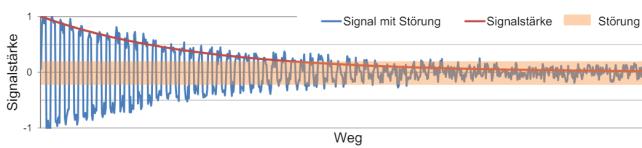
Funk- oder Licht-Signale sind elektromagnetische Wellen, die sich im Vakuum mit Lichtgeschwindigkeit $c_0 = 299'792'458$ ausbreiten. Die Vakuumgeschwindigkeit kann nicht überschritten werden.

$$C_{Medium} = 200'000 \text{ km/s} \approx 2/3 c_0$$

Signaldämpfung

Die Signaldämpfung bezeichnet die Leistungsabnahme eines Signals auf einer Übertragungsstrecke. Die Angabe der Signaldämpfung erfolgt in dB als logarithmische Verhältniszahl von Eingangsleistung P_1 zur Aufgangsleistung P_2 .

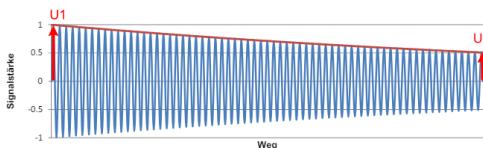
$$\text{Signaldämpfung [dB]} = 10 \cdot \log\left(\frac{P_1}{P_2}\right) = 10 \cdot \log\left(\frac{U_1}{U_2}\right)^2 = 20 \cdot \log\left(\frac{U_1}{U_2}\right)$$



Dämpfungsbelag

Für Übertragungsmedien ist die Dämpfung pro Distanz massgebend. Typischerweise in dB pro 100 m angegeben.

Berechnung Signaldämpfung/SNR



$$U1/U2 = 1/0.5 = 2, \text{ Signaldämpfung} = 20 \cdot \log(2) = 6 \text{ dB}$$

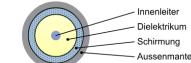
Kabeltypen

Overview

- Koaxialkabel: Geeignet für hochfrequente Signale
- Twinaxialkabel: Hoher Schutz (double koax)
- Twisted Pair (TP): Häufig im Einsatz (Shielded/Unshielded)
- Glasfaser: Hohe Bandbreite, Geringe Dämpfung, resistent

Koaxialkabel

Eignen sich für hochfrequente Signale, unempfindlich gegenüber elektromagnetischen Störungen, früher Standard für Netzwerke, aber teuer und mechanisch heikel



Paarsymmetrische Kabel (Twisted Pair)

Häufig im Einsatz, auch für breitbandige Datenübertragung nutzbar, Unterscheidung zwischen Shielded(STP)/Unshielded(UTP)

- Schirmeigenschaften
 - Drahtgeflecht: niederfrequente Einstreuungen
 - metallisch beschichtete Folien: hochfrequente Störungen
- Bezeichnungsschema ISO/IEC 11801
xx/yTP worin TP für Twisted Pair steht:

xx steht für die Gesamtschirmung:

U = ungeschirmt

F = Folienschirm

S = Geflechtschirm

SF = Schirm aus Geflecht und Folie

y steht für die Aderpaarschirmung:

U = ungeschirmt

F = Folienschirm

S = Geflechtschirm

Anfälliger auf Störungen (crosstalk), kapazitiv oder induktiv, Methoden zur Behebung:

- Kapazitiv: Komplementäres Signal, elektrisch leitenden Schirm
- Induktiv: Verdrillte Aderpaare

Lichtwellenleiter

Hohe Bandbreite, Geringe Dämpfung, Resistent, Dispersion als begrenzender Faktor

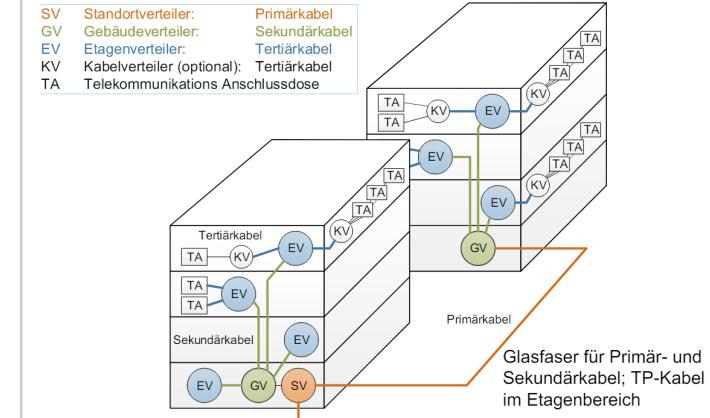
- Multimode: dicker Kern, günstiger, kleinere Datenraten und Übertragungsstrecken
 - Stufenfasern
 - Gradientenfasern
- Singlemode: dünner Kern, teuer!! aber funktionieren super

Grundprinzip optischer Fasern beruht auf der Totalreflexion und der Ausbreitung des Lichtes in bestimmten Moden

Key Takes

- Halbierung der Leistung entspricht ca. 3dB
- für SNR gleiche Formel wie Signaldämpfung, aber $P_1 = P_{Signal}$ und $P_2 = P_{Störung}$
- Entscheidend für die maximale Leitungslänge sind Dämpfungsbelag und SNR
- senkt man die Bitrate (Bit/s) können grössere Distanzen erreicht werden
- Die Bandbreite (Frequenz) ist in der Grafik abhängig zum Dämpfungsbelag.
- Die höheren Kabelkategorien brauchen, um höhere Dämpfung zu tolerieren, bessere Schirmungen, um das Übersprechen zu minimieren.

Strukturierte Gebäudeverkabelung nach ISO/IEC 11801



OSI Referenzmodell

Schichten, Protokolle und Dienste

Dienst sendet und empfängt bestätigte und unbestätigte Daten.

Klassifizierung von Diensten

Verbindungsorientiert

- Verbindungsaufbau nötig
- Informationen vom Empfänger - Optionen aushandeln
- Reihenfolge der Daten bleibt erhalten

Verbindungslos

- Jederzeit (send and forget)
- Ziel muss nicht bereit sein
- einfacher umzusetzen

Schicht hat die Aufgabe der darüberliegenden Schicht bestimmte Dienste zur Verfügung zu stellen. Die Schichten benötigen kein Wissen über die Realisierung der darunterliegenden Schicht.

Protokoll eine Sammlung von Nachrichten, Nachrichtenformaten und Regeln zu deren Austausch.

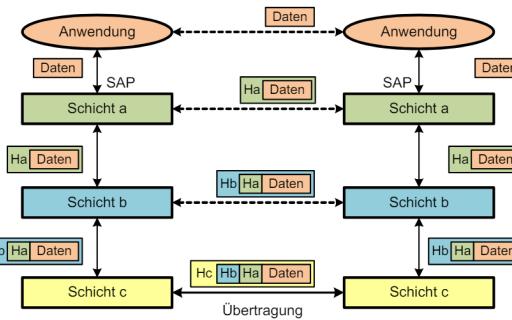
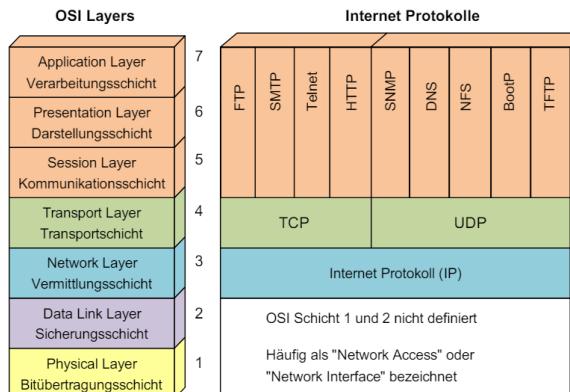
Im zwischenmenschlichen Bereich könnte man die Knigge als Protokoll bezeichnen. Sie legt einen gewissen «Verhaltens-Standard» nach welchem wir uns richten.

In der Technik ist ein Kommunikationsprotokoll eine Vereinbarung, die festlegt wie eine Datenübertragung zwischen Kommunikationspartnern abläuft.

Datenübertragung und Schichtenmodell

OSI Layers

- Anwendungsschichten 5-7
 - Lösen allgemeine Aufgaben
- Transportschichten
 - Teil des Betriebssystems und Treiber
 - «Socket»-Schnittstelle für eigene Anwendungsprotokolle



Key Takes

- Offene Systeme (im Gegensatz zu proprietären Systemen) basieren auf öffentlich verfügbaren Standards für Schnittstellen und Protokolle
- Referenzmodell, beschreibt Kommunikation zwischen räumlich entfernten Kommunikationspartnern
 - 7 Schichten (unabhängige Teilfunktionen)
 - Eine Schicht erfüllt, mit Hilfe der untergeordneten Schichten, bestimmte Aufgaben und bietet der übergeordneten Schicht über eine definierte Schnittstelle einen definierten Dienst an
 - Gleiche Schichten in verschiedenen Systemen kommunizieren nach den Regeln eines standardisierten Protokolls
- Die Kommunikation erfolgt logisch zwischen Protokollimplementierungen auf gleicher Schicht (Kommunikationsbeziehung), physikalisch wandern die Daten beim Sender im Stack «nach unten», werden über ein Medium übertragen, und wandern beim Empfänger wieder «nach oben»
- Dienste können als zuverlässig/unzuverlässig und verbindungsorientiert/verbindungslos klassifiziert werden

Physical Layer

Schicht 1: Bitübertragungsschicht



Funktionalität

Der Physical Layer sorgt für die ungesicherte Übertragung eines Bitstroms zwischen zwei Systemen

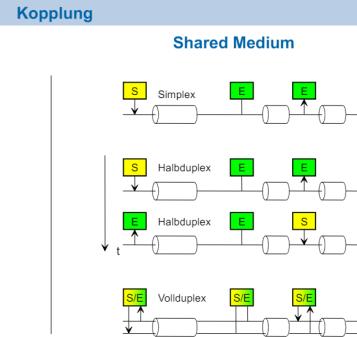
Die Standardisierung umfasst:

- Elektrische Eigenschaften (Signalform, Amplituden, Frequenzen etc.)
- Codierung (Abbildung der Daten auf elektrische Signale)
- Mechanische Eigenschaften (Stecker, Pinbelegung etc.)

Verschiedene Übertragungsmedien:

- Koaxialkabel, Twisted Pair, Lichtwellenleiter
- Radiowellen

Verkehrsbeziehung und Kopplung



Arten der Kommunikation (Verkehrsbeziehung)

- Simplex: Ein Kanal, eine Richtung
- Halbduplex: Ein Kanal, abwechselnweise in 2 Richtungen
- Voll duplex: Ein Kanal pro Richtung

Arten der Verbindungen (Kopplung)

- Punkt-Punkt: Direkte Verbindung zweier Kommunikationspartner
- Shared Medium: Mehrere Partner verwenden das gleiche Medium

Übertragungsverfahren: Parallel und Seriell

Parallel vs Seriell

- Parallele Übertragung: mehrere Bits gleichzeitig über mehrere Leitungen
- Serielle Übertragung (dominierend): einzelne Bits zeitlich gestaffelt

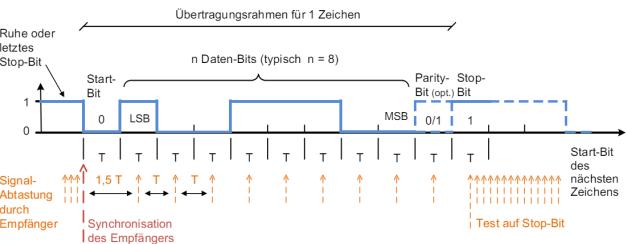
Bei der seriellen Übertragung wird weiter unterschieden zwischen seriell synchroner und seriell asynchroner Übertragung

Serielle asynchron Übertragung

Zwischen Sender und Empfänger werden folgende Abmachungen benötigt:

- Bitrate
- Anzahl Datenbits (Typisch 1 Byte)
- Anzahl Stoppbits (Typisch 1 Bit)

Taktrückgewinnung ist möglich



Welcher Wert / welches Zeichen wird hier übertragen?

- Empfänger wird 1001 1100 – LSB first -> 0011 1001 (binär); 0x39 (hex); ASCII Code 57 = «9»

Was ist die Genauigkeitsanforderung an die Takte von Sender und Empfänger (geometrisch ausgedrückt)?

- Letzte Abtastung muss noch im Zeitfenster liegen (Stop-Bit bei einem Stop-Bit); also $\frac{1}{2}T$ auf $9\frac{1}{2}T$

Clock Drift

Ethernet hat eine maximale Framegrösse von 1'500 Bytes.

- Gemäss Standard müssen die Oszillatoren eine Genauigkeit von ± 50 ppm haben.
- 50 ppm (parts per million) entspricht einem Fehler von 0.00005
- Der Worst-Case ist, dass der Sender einen Fehler von -50 ppm und der Empfänger einen Fehler von +50 ppm aufweist (oder umgekehrt)

Können in diesem Fall die Daten sicher abgetastet werden?

- 1'500 Bytes = 12'000 Bit; 100ppm Differenz zwischen Sender und Empfänger = $100 * 10^{-6} = 1 * 10^{-4}$
- Pro Bit entsteht so ein Fehler von 10-4 Bit-Zeiten TBit.
- 1'500 Bytes sind $12'000 = 1.2 * 10^4$ Bit
- Die Abweichung ist somit $1.2 * 10^4$ Bit * 10-4 TBit / Bit = 1.2 Tbit
- Eine fehlerfreie Abtastung ist nicht mehr möglich (ohne weitere Massnahmen)

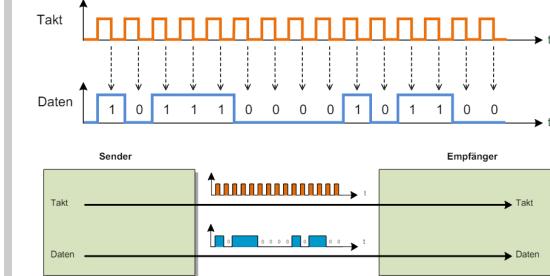
Serielle synchron Übertragung

Bei der synchronen Übertragung arbeitet der Empfänger mit dem gleichen Takt wie der Sender

- Keine Start- und Stoppbits benötigt
- Der Takt muss zusätzlich übertragen werden

Die Übertragung des Takts erfolgt über ein Codierungsverfahren oder eine zusätzliche Leitung.

Es ist die Aufgabe vom Data Link Layer die Grenzen der einzelnen Bytes zu ermitteln (Preamble, etc.)

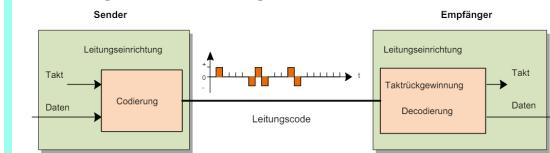


Welches Bit vom obigen Diagramm trifft zuerst beim Empfänger ein (1/0)? "1"

Vorsicht, wenn Weg und Zeit im selben Bild gezeichnet sind.

Synchrone Übertragung ohne separate Taktleitung

Geeignete Codierverfahren erlauben den Takt zusammen mit dem Datensignal zu übertragen



Unter Codierung versteht man hier die Umsetzung der Einsen und Nullen auf eine physikalische Grösse

- Vorteil: Es wird nur eine Leitung benötigt
- Nachteil: Zusätzlich 2 x Leitungseinrichtung

Leitungscodes

Leitungscodes und Taktrückgewinnung

Mittels Leitungscode ist es dem Empfänger möglich den Takt heraus zu extrahieren (sonst bräuchte er eine 2te Leitung für den Takt).

- Siehe Manchester Code.
- Regelmässige Zustandsänderungen auf der Übertragungsstrecke.

Anforderungen

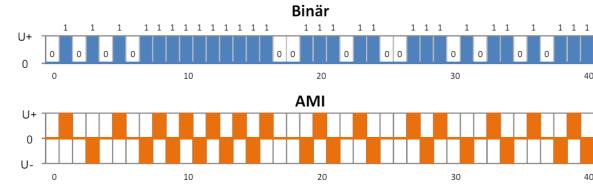
- die physikalisch vorhandene Bandbreite effizient nutzen
- Taktrückgewinnung erlauben, um eine separate Taktleitung einzusparen
- möglichst gleichspannungsfrei sein, um Sender und Empfänger mit Übertragern (Signaltransformatoren, Magnetics) galvanisch trennen zu können.

Wie könnte man Gleichspannungsfreiheit und galvanische Isolation in einem Schritt erreichen?

- Z.B. durch den Einsatz von Lichtwellenleitern

Gleichspannungsfreiheit

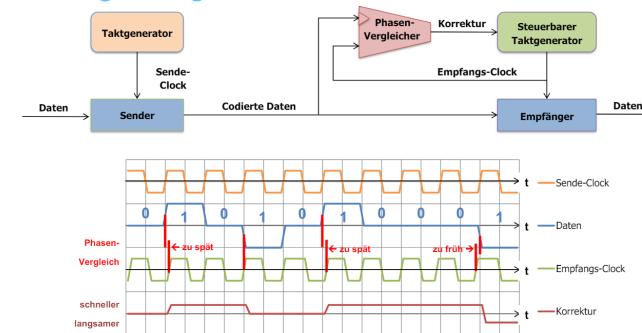
Beispiel: 3-wertiger AMI-Code (Alternate Mark Inversion)



Nachteile dieses Leitungscodes:

- Auf der Übertragungsstrecke drei Zustände benötigt → rein binäre Medien genügen nicht
- Eignet sich dieser Code gut für die Taktrückgewinnung im Empfänger?
- Bei einer längeren Folge von 0 in den Daten ist keine Taktrückgewinnung mehr möglich

Taktrückgewinnung



Datenrate, Bandbreite, Bandrate

Leitungssymbol Zu einem gewissen Zeitpunkt übertragenes physikalisches Signal, das mit einer bestimmten Rate (Symbolrate) seinen Wert (Eigenschaften, z.B. Amplitude, Frequenz, Phasenlage) verändert.

Wichtige Kenngrößen

- Bandbreite B – Einheit Hertz (Hz)
 - Eigenschaft des Übertragungskanals und durch das Medium begrenzt
- Symbolrate f_s – Einheit Baud (Bd)
 - Anzahl der Symbole pro Zeit. Limitiert durch die Bandbreite ($\leq 2B$) (Nyquist)
- Bitrate R – Einheit Bit/s (bps)
 - Produkt von Symbolrate und mittlerem Informationsgehalt der Symbole (Hartley)
- Kanalkapazität C – Einheit Bit/s (bps)
 - Berücksichtigt für einen realen Kanal das Signal-zu-Rausch Leistungsverhältnis S/N (Shannon)
- In der Kommunikation stehen k, M, G etc. SI-konform für die exakten Zehnerpotenzen:
 - $k\text{Bit} = 10^3 \text{ Bit}$, $M\text{Bit} = 10^6 \text{ Bit}$, $G\text{Bit} = 10^9 \text{ Bit}$
- Bitrate/Datenübertragungsrate/Durchsatz werden synonym verwendet

Datenübertragungsrate

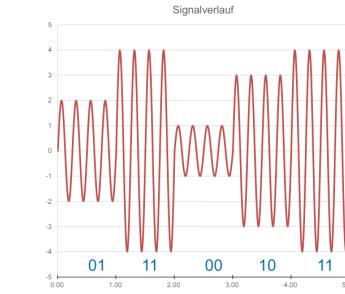
Die maximale Symbolrate f_s (Baud) ist gleich der doppelten Bandbreite B (Hz) des Übertragungskanals.

$$f_s = 2B$$

Beispiel Amplitude Shift Keying (ASK-4)

4-wertige Symbole, die sich nur in der Amplitude unterscheiden

- Baudrate: 1 kBaud
- Bit pro Symbol: $ld(4) = 2$
- Bitrate: 2 kBit/s
- Zusatzfrage
Trägerfrequenz: 4 kHz



Maximal erreichbare Bitrate

Maximale Bitrate $R[\text{bit}/\text{s}]$

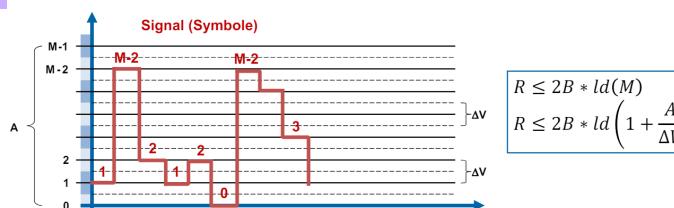
$$R \leq 2B \cdot \log_2(M)$$

Unterscheidbare Signalzustände

$$M = 1 + \frac{A}{\Delta V}$$

A = Max. Grösse des Signals

V = Ungenauigkeit des Empfängers



Kanalkapazität

$$C_s = B \cdot Id\left(1 + \frac{S}{N}\right)$$

S: Signalleistung

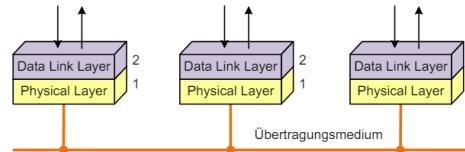
N: Rauschleistung

Key Takes

- Die physikalische Schicht befasst sich mit der Umwandlung physikalischer Signale (elektrisch, optisch) in einen Bitstrom und umgekehrt.
- Verkehrsbeziehung (Simplex/Duplex), Kopplung (Punkt-Punkt oder Shared Medium) und Übertragungsverfahren (synchron/asynchron) sind bestimmende Eigenschaften.
- Die Leitungscodierung legt fest, wie genau diese Umsetzung erfolgt. Wichtige Anforderungen sind Gleichspannungsfreiheit und Taktrückgewinnung.

Data Link Layer

Schicht 2: Sicherungsschicht

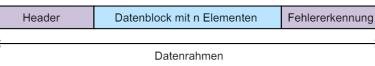


Aufgaben

- Realisieren einer zuverlässigen (fehlerfreien) Verbindung zwischen direkt miteinander verbundenen Systemen
- Framing (Rahmenbildung/-erkennung)
 - Senderichtung: Einpacken der zu sendenden Nutzdaten in Datenrahmen (Frames)
 - Empfangsrichtung: Erkennung und Auspacken der Datenblöcke aus empfangenen Frames
- Fluss-Steuerung (Flow Control)
 - «langsam» Empfänger kann «schnellen» Sender bremsen
- Zusätzlich bei mehreren Teilnehmern:
 - Adressierung der Teilnehmer
 - Medium Zugriff (Media Access)
 - Welche Station darf wann senden

Framing

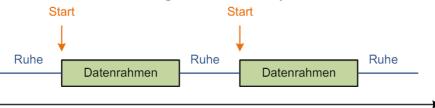
Frame besteht meist aus einem Header mit der Anzahl der Elemente im Datenblock, dem Datenblock und einem Code für die Fehlererkennung



Asynchron

Keine Daten → Nichts wird gesendet (Pause zwischen Frames)

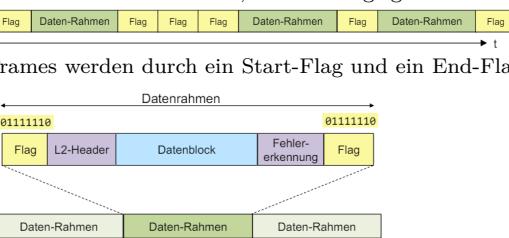
- Zu Beginn eines Frames wird ein Start-Bit gesendet
- Prüfbits am Ende eines Frames!
- Frame-Grenze gibt auch Byte-Grenze



Synchron

Frames werden ohne Unterbruch gesendet (kontinuierlicher Bitstrom auf Physical Layer)

- Stehen keine Daten an, werden Flags gesendet

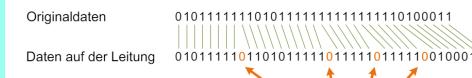


Maskierung von Sonderzeichen (Flags) nötig!

Bitstopfen

Wird verwendet um ein Bit-Muster zu garantieren.

- Sender fügt im Datenstrom nach 5 Einsen immer eine Null ein
- Empfänger wirft nach 5 Einsen immer ein Bit weg
- Somit gibt es (ausser bei Flags) die Bitfolge 0111110



Fehlererkennung/-Korrektur

Fehlerwahrscheinlichkeit

BER (Bit Error Ratio):

- Eigenschaft des Physical Layers
- wird als Dezimalzahl ausgedrückt: $BER = 0.5 \rightarrow$ jedes 2. Bit falsch

Weitere Definitionen:

- FER (Frame Error Ratio): Fehlerhaft empfangene Daten (Frames)
- RER (Residual Error Ratio): Unentdeckte, fehlerhaft empfangene Frames

Frame-Fehlerwahrscheinlichkeit Wie gross ist die Wahrscheinlichkeit, dass ein Frame der Länge N mindestens einen Bitfehler enthält? Für $BER = p_e \ll 1$ gilt: $(1 - p_e)^N \approx (1 - N \cdot p_e)$, also:

$$P_{\text{Fehler, Frame}} \approx N \cdot p_e (= FER)$$

Wahl der Framelänge

Die Wahl der optimalen Framegrösse ist ein Kompromiss zwischen Overhead und einer geringen Frame- und Restfehlerwahrscheinlichkeit. Sie wird von der Bitfehlerwahrscheinlichkeit, der Datenrate und Verzögerungen im System beeinflusst.

- Lange Frames:
 - Höhere Nutzdatenrate (höhere Netto-Bitrate, weniger Overhead)
 - Fehlerwahrscheinlichkeit wird grösser
 - Datenverlust bei einem Fehler wird grösser
 - Wahrscheinlichkeit eines unentdeckten Fehlers wird grösser
- Kurze Frames: Tiefe Nutzdatenrate, Zuverlässigkeit

$$\text{Framelänge} \quad \text{Nettobitrate} = \text{Bruttobitrate} \cdot \frac{\text{Nutzdaten}}{\text{Nutzdaten} + \text{Header}}$$

Datenraten

$$F_R = \frac{B}{8 \cdot (F_L + IFG)}$$

F_R = Framerate, B = Bitrate, F_L = Framelength

$$N = F_R \cdot P \cdot 8$$

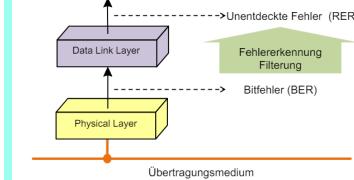
N = Nutzbitrate, P = Payload

Example or KR for actual exercises!!

Verfahren zur Fehlererkennung

Fehlererkennung

Die Zuverlässigkeit der Fehlererkennung ist abhängig von der Framegröße und gewähltem Verfahren

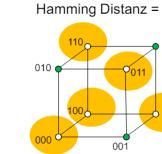


Standards IEEE 802 (LAN-Standards, z.B. Ethernet):

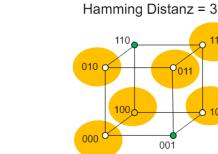
- max. $5 \cdot 10^{-14}$ unentdeckte Fehler pro Frame-Byte
- $BER p_e \leq 10^{-8}$
- CRC32 für Ethernet, mit Generatorpolynom

Hammingdistanz als Mass für Fehlerdetektion

Codes mit Hamming-Distanz ≤ 2 erlauben die Erkennung von Ein- oder Mehr-Bitfehlern



Erkennung eines einzelnen Bitfehlers



Erkennung von zwei Bitfehlern

Generelle Regel für Hamming-Distanz h : erlaubt Erkennung von $(h-1)$ Fehlern

$$e = h - 1$$

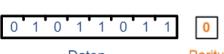
Fehlererkennung mit einfacher Parity

Ein (1) Prüfbit sichert ein Datenwort (typisch 1 Byte, auch 7 oder 9 Bits werden verwendet)

Even Parity

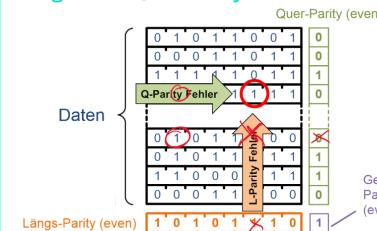


Odd Parity



- Even Parity: Anzahl 1er inkl. Parity-Bit ist gerade
 - Odd Parity: Anzahl 1-Bit inklusive Parity-Bit ist ungerade
- Even und Odd Parity sind gleichwertig

Längs- und Quer-Parity



Wie viele Fehler können korrigiert/erkannt werden?

- Korrigieren: 1 Bit-Fehler
- Erkennen: mind. 3 Bit-Fehler

Ethernet und LAN

Local Area Networks (LAN)

LAN

Räumlich kleines Netzwerk mit hoher Geschwindigkeit

- 10 m .. wenige km
- 100 Mbps .. 100 Gbit/s, typisch heute 1 Gbit/s

Verbindet Server, Workstations, PCs, Drucker, NAS, ...

- Auf Schicht 2 (direkte Kommunikation zwischen allen Stationen innerhalb des LAN)

Vorteile/Möglichkeiten:

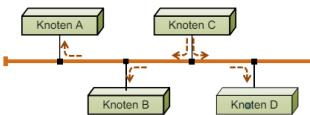
- Gemeinsamer Zugriff (Drucker)
- Datenaustausch (Directory/File Sharing)
- Direkte Kommunikation (VoIP, Gaming)
- Zugang zum Internet (über einen Router im LAN)

Topologien

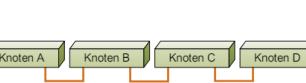
Bus

- Alle Stationen
 - sind passiv angeschlossen
 - horchen Leitung permanent ab
 - werden aktiv, wenn sie etwas senden wollen
- Keine festgelegte Ausbreitungsrichtung
- Empfänger erkennt anhand einer Adresse, ob die Daten für ihn relevant sind

Bus



Linie



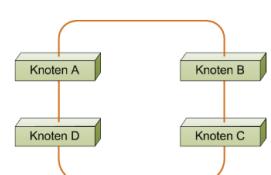
Linien

- Punkt-zu-Punkt Verbindungen zwischen benachbarten Knoten
- Alle Stationen müssen
 - Daten empfangen
 - Daten regenerieren
 - falls nötig weiterleiten
- Der Ausfall einer Station führt zur Segmentierung des LAN in zwei Teile

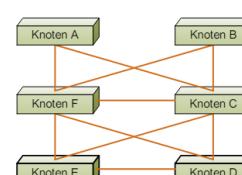
Ring

- Benötigt Verfahren zur Verhinderung von «endlosem Kreisverkehr»
- Gewisse Redundanz: beim Ausfall einer Station kann immer noch jede Station erreicht werden

Ring



Vermascht (teilweise oder komplett)



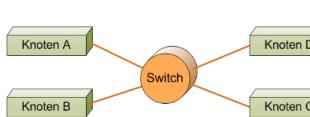
Vermascht

- Weitere Erhöhung der Redundanz:
 - Ausfall einer oder eventuell auch mehrerer Stationen oder Verbindungen kann toleriert werden
 - Zusätzliche Kosten und Aufwand, um mehrfache Lieferung von Daten zu verhindern

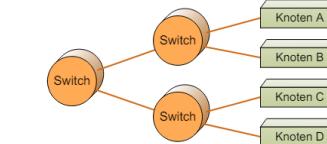
Stern

- Jede Station an zentralen Verteiler (Switch/Bridge) angeschlossen
- Verteiler entkoppelt Knoten elektrisch und macht LAN weniger störungsanfällig
- Verteiler sendet Daten, die er von einer Station erhält, an die anderen Knoten weiter

Stern



Baum



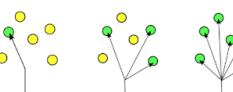
Baum

- Hierarchische Erweiterung der Sterntopologie
- Intelligenten Switches ermöglichen einen Grossteil der Kommunikation „lokal“
 - zwischen A und B bzw. C und D
- Dadurch Verringerung der Last für die einzelnen Switches

Übertragung und Adressierung

Übertragungsarten

- In jedem Fall: genau 1 Sender
- Unicast:
 - Genau ein, klar spezifizierter Empfänger
 - Frame trägt die Adresse dieses Empfängers
 - Analogie: Briefpost
 - Multicast:
 - Eine Gruppe von Empfängern
 - Frame trägt die Multicast-Adresse der Gruppe
 - Analogie: Mailing-Liste
 - Broadcast:
 - An alle Knoten im LAN gerichtet
 - Frame trägt die Broadcast-Adresse des LAN
 - Analogie: Radio-Sendestation



Diese Begriffe werden nicht nur im LAN, sondern auch allgemein in Netzwerken verwendet

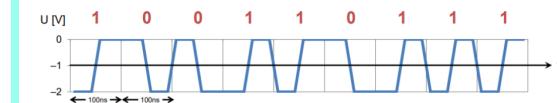
Adressierung in LANs

- Jede Station kann Daten von jeder anderen Station direkt empfangen
- Um zu erkennen, ob empfangene Daten für die eigene Station bestimmt sind, und wer der Absender ist, müssen die Adressen im LAN eindeutig sein.
- IEEE MAC Adressen
 - werden nicht konfiguriert
 - sind fix einem Interface des Gerätes zugeordnet
 - bestehen aus 6 Bytes
 - Darstellung hexadezimal 1A-2B-3C-4E-5F-67
- Geräte sind möglicherweise mobil und wechseln zwischen LANs, oder LANs werden direkt verbunden → Leitungscode

Manchester Leitungscode (10BASE-T)

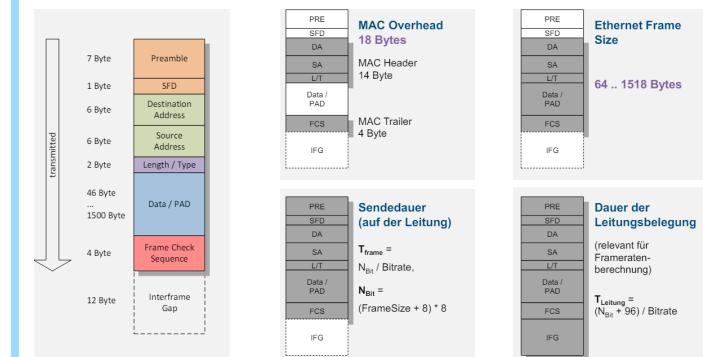
Es wird ein Manchester-Code gesetzt:

- 1 positive Flanke, 0 negative Flanke
- Bei jedem Bit gibt es einen Signalwechsel
- Erlaubt die Taktrückgewinnung auf einfache Weise
- Bandbreite von 10 MHz benötigt (also das doppelte des theoretischen Minimums)



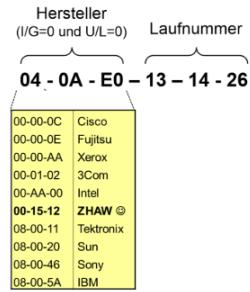
Ethernet Format Length/Type (2 Bytes)

- Fall 1: Länge von DATA ohne PAD (≤ 1500)
- Fall 2: Typ von Data = Protokoll der nächsten Schicht (≤ 1536)
- Data / Padding (46 – 1500 Bytes)
 - Enthält die eigentlichen Datenbytes
 - Bei weniger als 46 Bytes wird mit PAD Bytes abgefüllt
- Frame Check Sequence, FCS (4 Bytes)
 - IEEE CRC-32 Algorithmus
 - Interframe Gap, IFG (12 Bytes)
 - «Zwangspause» zwischen aufeinanderfolgenden Frames
- Ist NICHT Teil des Ethernet Frames



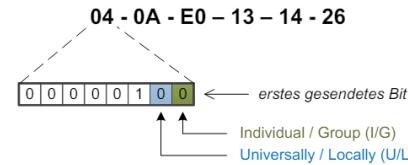
IEEE MAC Adressen

- Registrierung bei IEEE:
- 3-Byte «OUI» identifiziert Hersteller
- 3-Byte Laufnummer durch Hersteller verwaltet

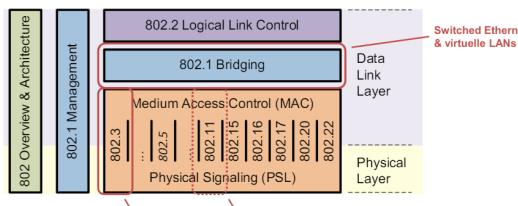


Die ersten beiden Bits des ersten Adress-Bytes klassifizieren die MAC Adresse:

- Individual / Group Bit
 - 0 = individual address
 - 1 = group address
- Universally / Locally Bit
 - 0 = universally administrated address
 - 1 = locally administrated address



IEEE 802 Standards



Ethernet Frame Format

Sie senden ein Ethernet Frame über eine 100BASE-TX Schnittstelle und beobachten auf dem Kabel folgende Bit-Sequenz:

10101010 10101010 10101010 10101010 10101010 10101010
 10101010 10101011 00001000 00000000 01011010 11100011
 10011111 10000110 ...

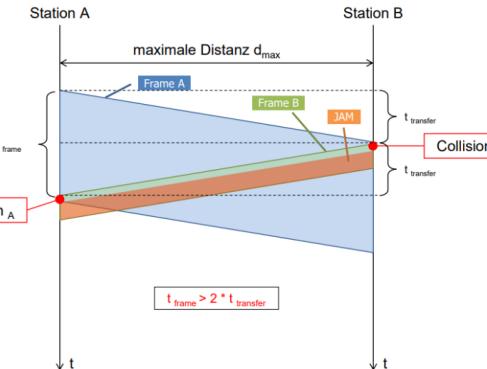
Wie lautet die MAC-Adresse (in hexadezimaler Darstellung) des Empfängers des Frames und wer ist der Hersteller der Ethernet-Karte dieses Empfängers? (Hinweis: von den einzelnen Bytes des Frames wird zuerst das LSB und am Schluss das MSB übertragen!)

- Zuerst 7 Bytes Präambel (10101010), dann 1 Byte SFD (10101011)
- 6 Bytes Destination Address: 00001000 (=08) 00000000 (=00) 01011010 (=5A) 11000111 (=C7) 11111001 (=F9) 01100001 (=61)
- MAC-Adresse: 08-00-5A-C7-F9-61, Hersteller (08-00-5A) IBM

Kollisionen

- Bei Überlagerungen von Signalen. (Ergänzung 2 Schicht 2)
- Bspw. zwei Frames kommen gleichzeitig im Hub (Schicht 1) an (Minimaler Switch (Schicht 2) kann dies nicht passieren).
- Vollduplex: Keine Kollision da keine Spannungserhöhung (kann man prüfen).
- Hub / Repeater erkennt Kollisionen wenn gleichzeitig von mehreren Ports Frames empfangen werden.

Kollisionserkennung können durch Überlagerung von Signalen entstehen. Kollisionen müssen erkannt werden!



Bedingungen für Kollisionserkennung:

- Ohne Repeater: $t_{frame} > 2 \cdot t_{transfer}$
- Mit Repeater: $t_{frame} > 2 \cdot (\sum t_{transfer} + \sum t_{forwarding})$

Maximale Ausdehnung eines Segments:

$$t_{frame} = \frac{\text{Framesize}_{min}}{\text{Bitrate}}, t_{transfer} = \frac{d_{max}}{C_{Medium}}$$

Ein Knoten kann Kollisionen nur lokal erkennen, solange er selbst am Senden ist

$$d_{max} < \frac{1}{2} \cdot \frac{\text{Framesize}_{min}}{\text{Bitrate}} \cdot C_{Medium}, d_{max} < \frac{1}{2} \cdot \frac{576\text{Bit}}{10 \cdot 10^6 \cdot \text{Bit/s}}$$

Switched LAN und Ethernet

Switch/Brigde

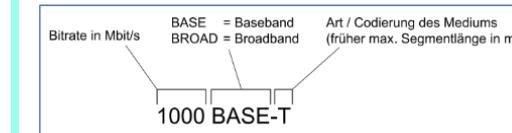
- Verwenden «Filtering Database».
- Switch lernt nur die Senderadressen nicht den Empfänger.
- Unbenutzte Einträge werden nach einer gewissen Zeit gelöscht.
- Port Mirroring möglich

Ethernet Geräte (Network Gear)

Repeater and Collision Domain

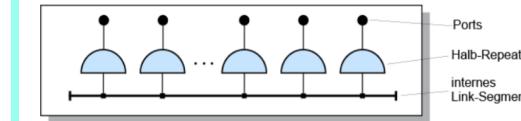
Eine Collision Domain ist ein Teilbereich eines LANs, in dem die Frames der Stationen miteinander kollidieren können.

- Erkennen von Kollisionen
 - Halbduplex Collision Detection Unit
 - Vollduplex Keine Kollisionen
- Shared Medium Ethernet
 - Carrier Sense Multiple Access with Collision Detection (CS-MA/CD)
- Normen für CSMA/CD
 - Verbilligung (Thick Ethernet → Thin Ethernet)
 - Vereinfachung (Koaxial → Twisted Pair)
 - Leistungssteigerung (10 → 100 … 100'000 Mbit/s)



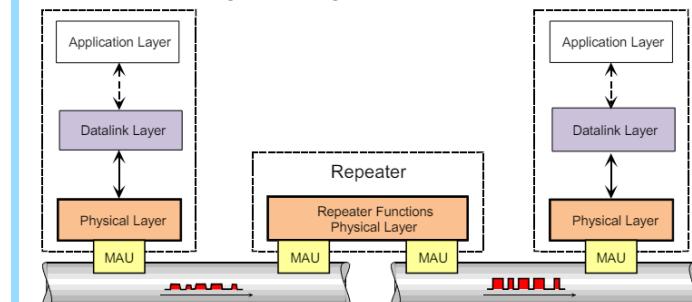
Repeater/Hub

- Ankommendes Signal wird an alle anderen Ports weitergeleitet, regeneriert und ausgesendet.



Repeater/Hubs im OSI Modell

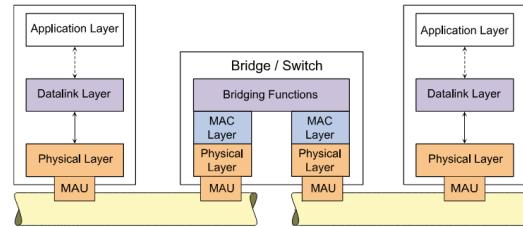
- Verstärkt ankommende Signal auf einem Port und leitet sie «in bester Qualität» weiter
 - Aufgetretene Bitfehler werden ebenfalls weitergeleitet / keine Fehlererkennung
 - Signalpegel, Signalfanken etc. sind regeneriert
- Medienkonverter (elektrisch-optisch oder COAX-Twisted Pair) sind funktionell gleichwertig



Switch im OSI Modell

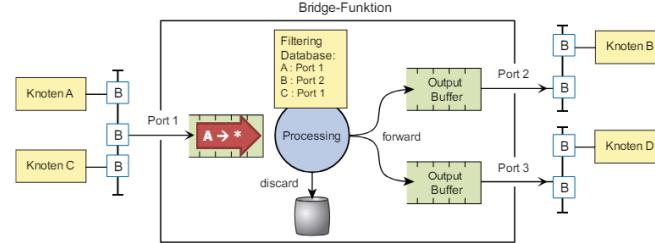
Switch arbeitet auf der Schicht 2: IEEE nennt einen Layer-2 Switch eine Bridge

- Prüft Checksumme und kann Layer-2 Adressen auswerten
- hat bereits seit längerem HUBs verdrängt, keine Kostenvorteile mehr



Filtering Database

Switches verbinden LAN-Segmente



Transparente Switches

- Bridges sollen für Endgeräte unsichtbar sein
- Sicht Endgerät: A und B sind direkt verbunden
- Konsequenz?
 - Bridges dürfen Verkehr nur dann filtern, wenn sicher kein potentieller Empfänger ausgeschlossen wird
 - «Flooding» für Broadcast und Multicast Zieladressen
 - Unicast Verkehr kann gefiltert werden, wenn bekannt ist, über welchen Port das Ziel erreicht werden kann
- Address Learning
 - Aufbau und Update der Filtering Database durch Verkehrsbeobachtung (Absenderadresse)
 - Alte Einträge werden gelöscht, wenn kein Verkehr vom Absender mehr beobachtet wird

Bridges

Bridges verfügen über einen Mechanismus zum Erlernen von Adressen. Eine Bridge hört den Verkehr von allen Ports ab und merkt sich die Sender-Adressen aus den empfangenen Frames in der sogenannten «Filtering Database». Diese beinhaltet für jede bekannte Mac-Adresse das Bridge-Port, über welches der zugehörige Knoten erreichbar ist. Unbenutzte Einträge in der Filtering Database werden nach einer gewissen Zeit automatisch gelöscht.

Diese Verarbeitung benötigt etwas Zeit, ist aber dennoch vorteilhaft, da das Paket nur an die richtige Collision Domain geschickt wird.

Multi-Port-Bridges

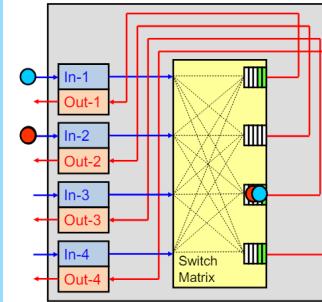
- Daten werden ausschließlich an den richtigen Port weitergeleitet.
- Standard-Komponente zur Kopplung von Segmenten
- Werden als Ethernet-Switch bezeichnet

Broadcast and Collision Domain

Eine Collision Domain (CD) besteht aus mit Repeatern zusammen geschlossenen Segmenten.

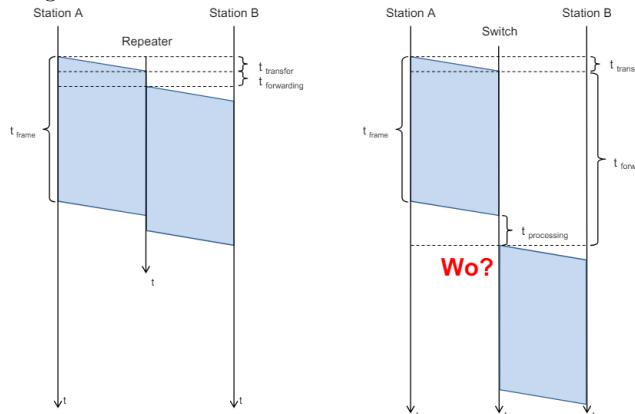
- Max. halb so lange wie die Ausdehnung des kürzesten Frames
- Ein virtuelle LAN bildet eine Broadcast Domain. Das heisst die Grenzen für die Verteilung der Broadcast-Frames.

Bridges: Arbeitsweise eines 4-Port Switches



Weg/Zeit-Diagramm für das Senden eines Frames

Gesamtübertragungszeit (Latenz): $t_{frame} + t_{transfer}$
 $t_{forwarding}$ kann verlängert werden um eine Verarbeitungszeit zu ermöglichen



Übersicht Netzwerkgeräte im LAN

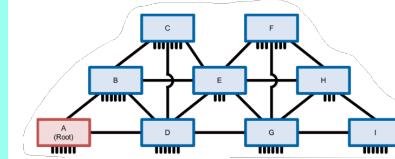
OSI Schicht	2-Port	Multi-Port
Data Link Layer	Dual-Port Switch (*) (IEEE: Bridge)	Ethernet Switch (IEEE: Multiport-Bridge)
Physical Layer	Repeater	Hub

Redundanz (Spanning Tree)

Spanning Tree Algorithmus

Redundante Pfade schaffen Probleme!

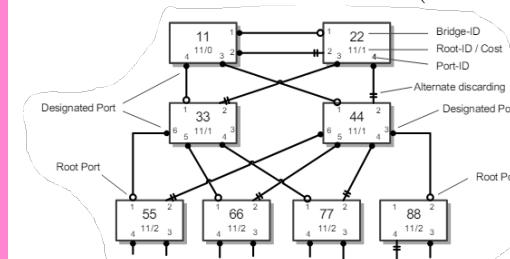
- Ziel: Alle Segmente in einer loop-freien Topologie verbinden
- Idee:
 - Root-Bridge auswählen (willkürliche, aber eindeutige Wahl)
 - * Auswahl ist vom Bridge-Identifier abhängig
 - * Bridge Identifier besteht aus einer wählbaren Priorität und der MAC Adresse
- Ausgehend von der Root einen Baum aufbauen
- Redundante Pfade sperren
- alle Knoten werden genau einmal verbunden



E wäre theoretisch besser geeignet als Root-Bridge

Spanning Tree Algorithmus Initialisierung

- Alle Ports für Nutzdaten blockiert
- Annahme: «Ich bin Root»
- Austausch BPDUs mit Nachbarn
- Aufbau des Spanning Tree
 - Aufdatieren der Info zu Root (kleinste ID) und Pfadkosten zu dieser
 - Austausch aufdatierter BPDUs bis Konvergenz
- Setzen der Port Roles
 - Freigeben für Nutzdaten von
 - Root-Ports (Empfang der «besten» BPDU)
 - Designated-Ports (Gegenstück zu Root-Ports)
 - Alle anderen Ports bleiben blockiert (Discarding)



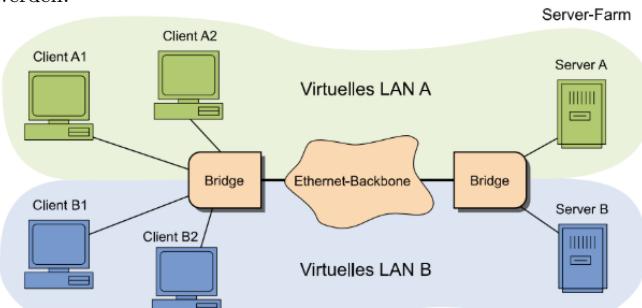
BPDUs (Bridge Protocol Data Units) beinhalten:

Root-ID (aus lokaler Sicht):	8 Byte
Root-Cost (aus lokaler Sicht):	2 Byte
Bridge-ID («Ich»):	8 Byte
Port-ID (Sende-Port):	2 Byte

Virtuelle LANs und Quality of Service

VLAN Mithilfe von virtuellen LAN kann ein grosses Netz in unabhängige logische Netze aufgeteilt werden. Jedes Switch-Port kann einem beliebigen VLAN zugeordnet werden.

Trunk Links sind Teil von mehreren VLANs. Auf den Trunk Links müssen Frames der verschiedenen VLANs eindeutig gekennzeichnet werden!



Trunk = Tagged, Access = Untagged

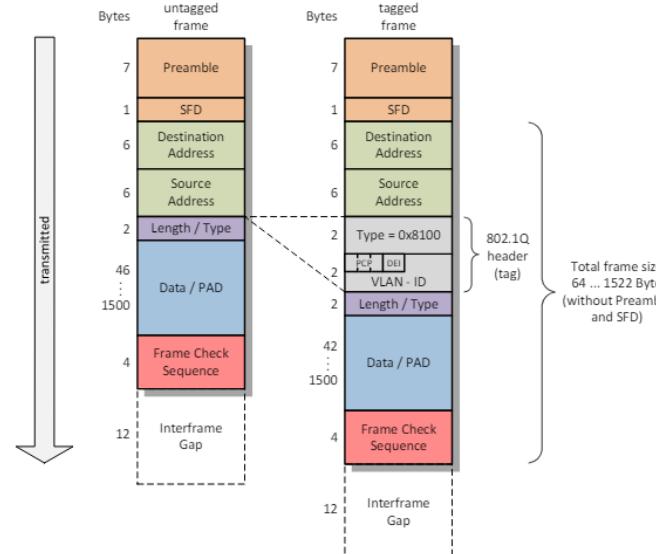
VLAN Tag

- VLAN-ID im VLAN-Tag wird zur Zuordnung verwendet
- Priority Code Point ermöglicht die Priorisierung gewisser Applikationen
- Discard Eligibility Indicator 0 → Frame wird bei Engpässen zuerst verworfen
- VLAN Tagging erfolgt oft beim Eintritt/Austritt ins Netz
- Vorteile:
 - Transparent (unsichtbar) für Endgeräte
 - VLAN Konfiguration nur im Netz
 - * Einfache zentrale Konfiguration
 - * Einfaches Anpassen der Konfiguration

VLAN Tagging

Erweiterung des Ethernet Headers durch einen VLAN-Tag

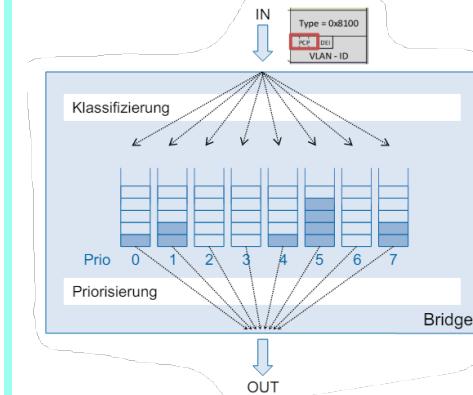
- Der Type 0x8100 bedeutet, dass das Frame «getagged» ist
- Ein 12 Bit Identifier (VLAN-ID, VID) besagt, welchem VLAN das Frame angehört
- Die 3 Bit des Priority Code Point (PCP) erlauben, das Frame mit einer Priorität zu versenden
- Mit dem Drop Eligibility Indicator (DEI) werden Frames markiert, die bei Überlastsituationen zuerst verworfen werden sollen
- Die maximalen Nutzdatenlänge bleibt erhalten, der Ethernet Frame wird 4 Bytes länger
- VLANs können transparent eingesetzt werden



Quality of Service (QoS)

Quality of Service: Priorisierung aufgrund der PCP-Werte (0 .. 7)

- Erlaubt Unterstützung verschiedener Verkehrsklassen im Layer 2, wie
 - Voice (< 10 ms Latenz)
 - Video (< 100 ms Latenz)
 - Best Effort Daten
 - Background Daten
- Jeder Switch-Port verfügt über mehrere Ausgangs-Queues (bis zu 8)
 - Frames können sich im Switch überholen
 - Verschiedene Priorisierungs-Strategien möglich, im einfachsten Fall «Strict Priority»
 - * Prio 7 vor Prio 6 ... vor Prio 0



Switched LANs: Merkmale von Switches/Bridges

Anzahl Ports	Steckergöße ist im Extremfall die Limitierung
Adressabelle	Wieviele Stationen können im LAN existieren
Filterrate	Maximale Frames / s / Port (Empfangsrichtung)
Transferrate	Maximale Frames / s / Port (Senderrichtung)
Backplane / Fabric Kapazität	Maximaler Gesamt durchsatz zwischen allen Ports
Architektur	Store-and-Forward: Frame wird komplett empfangen und dann weitergeleitet Cut-Through: Frame wird schon nach Decodierung der Zieladresse weitergeleitet Adaptive Cut-Through: Leitet auch korrupte Frames weiter, in der Regel aber kein Problem
Konfigurierbarkeit	Unmanaged (keine Möglichkeit z.B. VLANs einzurichten) oder Managed (via Konsole oder Web Interface)
Energieverbrauch	Wird zunehmend wichtiger in Data Center Anwendungen

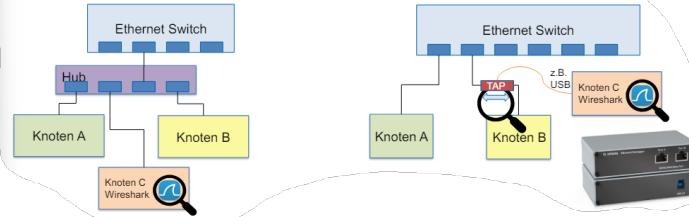
UNTERSCHIEDE BRIDGES UND SWITCHES

Switched LAN Monitoring

Hub/Multiport Reader

- Pro
 - Alle Daten sind auf allen Ports sichtbar
 - Con
 - Verfälscht die Situation völlig
 - Nur Half Duplex Betrieb für A und B möglich
- Tap/Probe
- Pro
 - Sehr detaillierte low-level Analyse möglich
 - Con
 - Kosten
 - Veränderung des Netzwerkes (Latenz)

Hub / Multiport Repeater

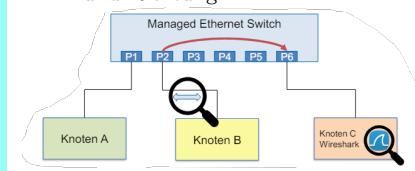


Port Mirroring Managed Switches unterstützen eine Vielzahl Diagnosefunktionen

- Statistik Zähler pro Port: Anzahl Rx und Tx Frames, FCS-Fehler, zu lange Frames, ...

Port-Mirroring leitet Daten zusätzlich auf einen anderen Port um

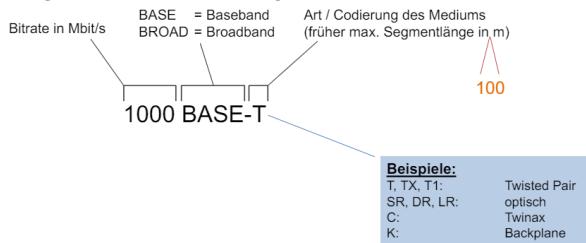
- Konfigurationsoptionen herstellerabhängig
 - Kompletter Port (Rx plus Tx), oder selektiv Port Nummer(n) und Richtung



Ethernet Systeme

Bezeichnungs-Schema

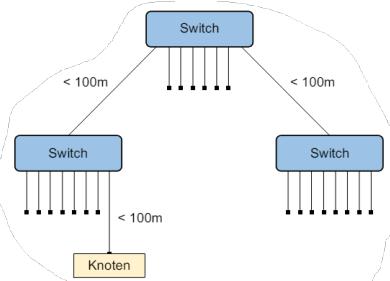
Jede im Standard IEEE 802.3 definierte Technologie hat einen dreiteiligen Namen, der ihre Eigenschaften beschreibt



10BASE... heisst 10Mbps, 10GBASE heisst 10Gbps, usw.

Topologie aller Twisted-Pair-Ethernet

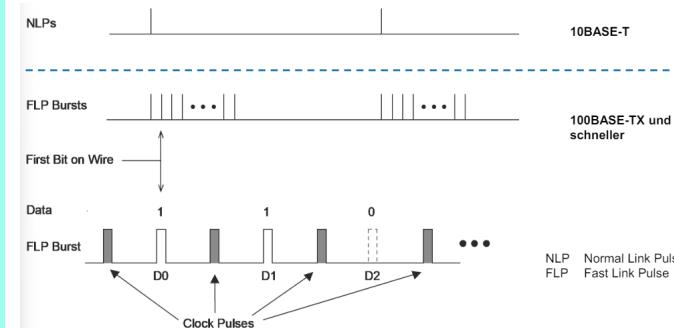
- Stern-Topologie mit mehreren Stationen an einem Ethernet Switch
- Mehrere Switches erweitern Topologie zu einem Baum
- Für Twisted Pair (10 Mbit/s .. 10 Gbit/s) gilt in der Regel:
 - 4 verdrillte Adernpaare (für < 1 Gbit/s auch 2)
 - Maximale Segmentlänge 100m
 - RJ45 Stecker
- Allenfalls vermascht für Redundanz



Autonegation

- Konzept
 - Ermittlung der besten Betriebsart durch Austausch der Leistungsmerkmale zweier Netzwerkkomponenten.
 - beruht auf Fast Link Pulses
- Link Pulses
 - NLP = Link Presence Detection
 - FLP = Autonegotiation, Autopolarity

	10BASE-T	100BASE-TX	1000BASE-T	10GBASE-T
Kabelkategorie	CAT3 - 16 MHz CAT5 - 100 MHz	CAT5 - 100 MHz CAT6 - 250 MHz	CAT5 - 100 MHz CAT6 - 250 MHz CAT7 - 600 MHz CAT7A - 1000 MHz	CAT6 - 500 MHz CAT7/7A - 600/1000 MHz
Line Coding	Manchester 2 Adernpaare simplex	MLT-3, 4B5B 2 Adernpaare simplex	PAM-5, 8B/10B 4 Adernpaare duplex	PAM-16, 64B/65B, FEC 4 Adernpaare duplex
Baudrate	10 Mbaud	125 Mbaud	4 x 125 Mbaud	4 x 800 Mbaud
Link Pulses	NLP	FLP	FLP	FLP



Leitungscodierung

NRZI-Codierung (Non Return to Zero Inverted), kombiniert mit MLT-3

011001010010001000110



MLT-3 = Multi-Level Transmit - Ternary

GBASE-T

Name	Standard	Speed (Mbit/s)	# TP	Coding	Baud rate per lane (Mbd)	Bandwidth	Max distance (m)	Cable	Cable rating (MHz)
10GBASE-T	802.3an-2006	10000	4	64B65B PAM-16 128-DSQ	800	400	100	Cat 6a	500
5GBASE-T	802.3bz-2016	5000	4	64B65B PAM-16 128-DSQ	400	200	100	Cat 6	250
2.5GBASE-T	802.3bz-2016	2500	4	64B65B PAM-16 128-DSQ	200	100	100	Cat 5e	100

Twisted Pair Ethernet Evolution Übersicht

	10BASE-T IEEE 802.3i	100BASE-TX IEEE 802.3u	1000BASE-T IEEE 802.3ab	10GBASE-T IEEE 802.3an
Kabelkategorie (für 100m Link Distanz)	CAT3 B = 16 MHz CAT5 B = 100 MHz	CAT5 B = 100 MHz CAT6 B = 250 MHz	CAT5 B = 100 MHz CAT6 B = 250 MHz	CAT6 B = 500 MHz CAT7/7A B = 600/1000 MHz
Line Coding	Manchester 2 Adernpaare simplex	MLT-3 (synchron), 4B5B 2 Adernpaare simplex	PAM-5 (plus scrambling) 4 Adernpaare duplex	PAM-16, 64B/65B, FEC 4 Adernpaare duplex
Baudrate	10 Mbaud	125 Mbaud	4 x 125 Mbaud	4 x 800 Mbaud
Link Pulses	NLP (Link Presence Detection)	FLP (Autonegotiation, Autopolarity)	FLP (Autonegotiation, Autopolarity, Next Page)	FLP (Autonegotiation, Autopolarity, Next Page)
Erfolgreich durch kompatible Elemente	Frameformat Adressformat 100m Kabellänge 64 B Mindestlänge RJ45 Stecker	Frameformat Adressformat 100m Kabellänge 64 B Mindestlänge RJ45 Stecker	Frameformat Adressformat 100m Kabellänge 64 B Mindestlänge GG6/RJ45/TER4-Stecker	

Key Takings LAN/Ethernet Basics

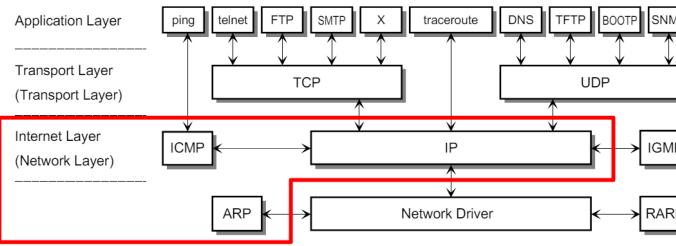
- LANs waren historisch räumlich kleine Netze, die verschiedene Geräte verbinden und in welchen Daten mit hoher Geschwindigkeit übertragen werden
- LAN-Topologien sind Bus, Linie, Ring, Stern, Baum und vermaschte Strukturen
- IEEE hat in den Normen 802 eine Reihe von Standards für LAN und MAN aufgestellt
- Alle Ethernet Varianten
 - Definieren die physikalische und Teile der Sicherungsschicht
 - verwenden das gleiche Frame-Format, das auch in allen später entwickelten Ethernet/802.3 -Varianten verwendet wird
 - MAC-Adressen der Länge 6 Bytes identifizieren Ethernet Geräte
- 10BASE-T ist bereits sehr alt, kann aber einfach beobachtet werden
 - Manchester Codierung, Unterstützung von Repeatern/Hubs
- Switches (Bridges) arbeiten transparent (unsichtbar) auf dem Data Link Layer und schliessen mehrere Segmente zu einem LAN zusammen
 - Bridges leiten (Address Learning) Frames nur dorthin weiter, wo sie empfangen werden müssen → Lastreduzierung und Erhöhung der effektiv nutzbaren Kapazität

Key Takings Ethernet-Technologie und -Systeme

- Mit gemanagten Switches/Bridges kann ein physikalisches LAN in mehrere virtuelle LANs (VLAN) mit separaten Broadcast Domains aufgeteilt und Prioritäten unterstützt werden
- Port Mirroring ist ein mögliches Verfahren zur Verkehrsbeobachtung und Fehlersuche in Switched Ethernet
- Redundanz wird ermöglicht
 - durch die „künstliche“ Reduktion der Topologie zu einer Baumstruktur durch Spanning Tree Algorithmen
- Kompatibilität (10)/100/1000BASE-T wird erreicht durch
 - Beibehaltung von Frame Format und Schnittstelle zwischen PHY und MAC
 - Autonegotiation mittels FLP bursts / NLP
- PHY Codierung ist unterschiedlich (Scrambled NRZ/MLT-3 mit 4 5 Codierung, ...)
 - Höhere Datenrate → höhere Ansprüche an die Signalverarbeitung und Algorithmen im PHY
- Massnahmen zur Reduktion von Maximaler Signalfrequenz und EMV Pegeln werden mit steigender Datenrate immer wichtiger
- Zur Zeit dominieren
 - Geswitchte 100BASE-TX oder 1000BASE-T Systeme, die oft gemeinsam in einem LAN betrieben werden
 - Geswitchter Vollduplex Betrieb (mikrosegmentiertes LAN)

Network Layer

Schicht 3: Internet



Die Netzwerkschicht

- verbindet verschiedene Netze
- ist in Endsystemen und den verbindenden Elementen (Routern) implementiert
- verbirgt die Besonderheiten der einzelnen Netze

Grundlegende Unterscheidung zwischen

- Forwarding (Weiterleitung der Daten)
 - Aufgrund von Routingtabellen
- Routing (Aufbau der Routingtabellen)
 - Durch statische Konfiguration oder
 - Dynamisch durch Routing-Protokolle

Was braucht es damit das Internet "funktioniert"?

- Information, zu welchem Knoten Daten geschickt werden sollen → Adressierung
- Geräte, die die verschiedenen LANs miteinander verbinden → Router
- Wie weiss ein Router, wohin die Daten weitergeleitet werden sollen? → Routing
- Was geschieht, wenn unterschiedliche LANs unterschiedliche Framegrößen unterstützen? → Fragmentierung
- Welche Helferlein fürs Internet gibt es? Was kann ich machen, wenn etwas nicht wie erwartet funktioniert? → Internet Control Message Protocol (ICMP)

Grundsätze des Internets

- Jedes Netzwerk soll für sich selbst funktionsfähig sein
- Die Kommunikation basiert auf „best effort“
- Die Verbindung der Netze erfolgt durch Black Boxes
- Keine zentrale Funktionssteuerung wird benötigt

Kommunikationsobjekte

Die vier Grundsätze führen zur Wahl eines paketvermittelnden Netzes auf der Grundlage von vier Layern (siehe OSI Modell)

- **(Application-)Message/Stream** Layer 5-7
- **(Transport-)Paket, Datagram** Layer 4
- **(IP-)Paket (früher Datagram)** Layer 3
- **(HW-specific) Frame** Layer 1-2

Eigenschaften des Internet Layers

- Die Basis bildet ein verbindungsloser Network Layer
 - Er hält das virtuelle Netz von Teilnetzen zusammen.
 - Er leitet IP-Pakete zwischen zwei beliebigen Hosts weiter; egal in welchen Teilnetzen sich die Hosts befinden.
- Kümmt sich nur um den Transport der IP-Pakete
 - Es ist nicht die Aufgabe des Internet Layers, den Verlust oder die fehlerhafte Übertragung von Paketen zu korrigieren.
 - Die Reihenfolge der Pakete kann ändern, falls Pakete auf verschiedenen Pfaden übertragen werden.
- Bei IP müssen die höheren Layer zusätzlich erwünschte Funktionalität übernehmen. Beispiele:
 - Fehlerfreie, komplekte Übertragung
 - Richtige Reihenfolge
 - Flusskontrolle

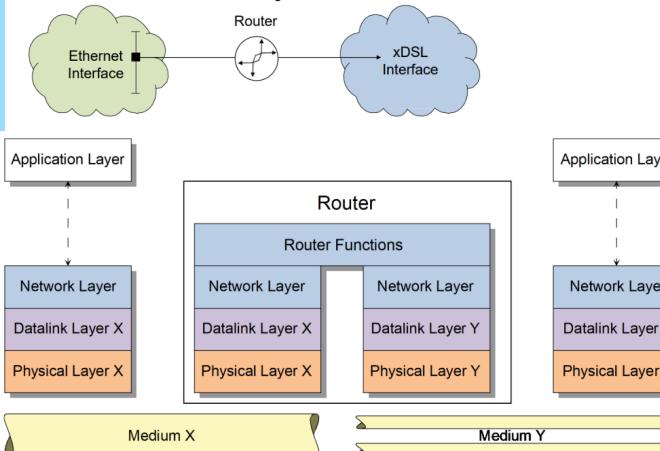
Netzwerk Applikationen und Protokolle

Routing

Router

Komponenten, die es erlauben Subnetze miteinander zu verbinden. Router haben eine ähnliche Funktion wie Bridges, allerdings arbeiten sie auf dem Network Layer.

- Router empfangen nur Pakete, die direkt an sie adressiert sind.
- Die Weiterleitung erfolgt anhand der Network Layer Adresse.
- Benutzen immer den optimalen Pfad.



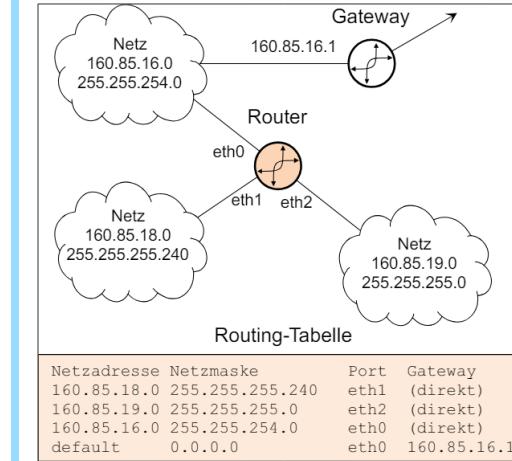
Routing and Forwarding

- **Routing:** Aufbau und Update der Routingtabellen in den Knoten
 - dazu müssen die Router den optimalen Pfad zu jedem Host kennen
 - In kleinen oder Teilnetzen: Statische Konfiguration
 - in grösseren Netzen: Dynamisch durch Routing-Protokolle die Topologie des Netzes ermitteln, daraus ideale Pfade bestimmen
 - **Forwarding:** Weiterleiten der Daten
 - Aufgrund von Routingtabellen Datenpakete weiterleiten
- Jeder Knoten auf dem Weg, einschliesslich dem Sender, wertet seine Routingtabelle aus und trifft Forwarding-Entscheidungen

Routing-Tabelle

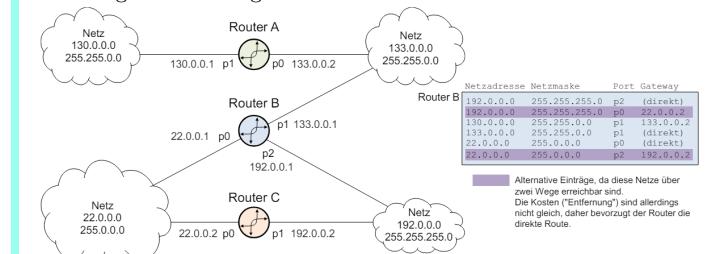
Enthält Informationen, wie jedes Netz (und damit jedes Interface) erreicht werden kann

- Für Weiterleitungsentscheidung notwendige Informationen
 - Welche Netze (Netzadresse, Subnetzmaske) gibt es? Wichtig für die Skalierbarkeit: Netz-Adressen, nicht Interface Adressen!
 - Über welches Interface sind diese Netze erreichbar?
 - Ist das Zielnetz erreicht, oder muss das Paket an einen nächsten Router (IP-Adresse) weitergegeben werden?
- Sortiert nach der Länge der Netzmasks
- Von oben nach unten durchsucht
- Verglichen werden die Netzadressen
- erster Eintrag der passt wird verwendet, default Eintrag am Schluss passt immer



Flaches Routing

- Router kennt explizite Wege zu jedem einzelnen Zielnetz
 - Pakete an unbekannte Netze werden verworfen
- Redundanz möglich durch Speichern mehrerer Wege ins gleichen Netz
 - Wege müssen nicht gleich gut sein
- Einsatz in stark vermaschten Netzen oder im zentralen Bereich (Backbone)
- Sehr grosse Routing-Tabellen



Flaches Routing Übung

Was geschieht mit dem IP-Paket?

- Kein Unterbruch?

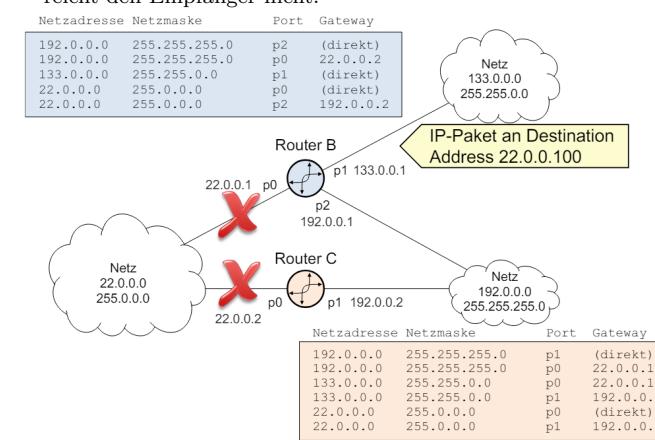
Es wird nach gemäss dem 4. Eintrag der Routingtabelle von Router B an p0 weitergeleitet

- Unterbruch von p0 / Router B ?

Es wird gemäss Eintrag 5 in der Routingtabelle von Router B an p2 weitergeleitet.

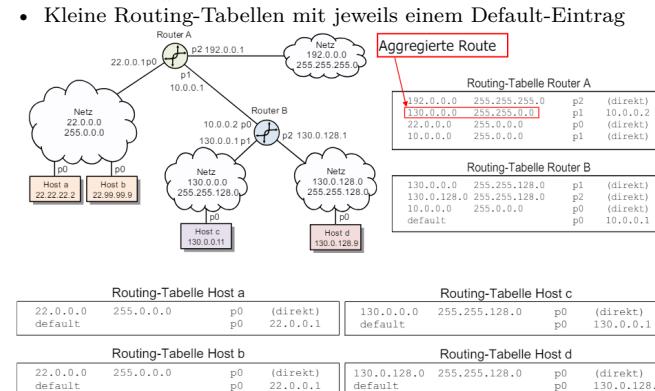
- zusätzlicher Unterbruch p0 / Router C ?

Router C kann das IP-Paket nicht weiterleiten, es IP-Paket erreicht den Empfänger nicht.



Hierarchisches Routing (Default)

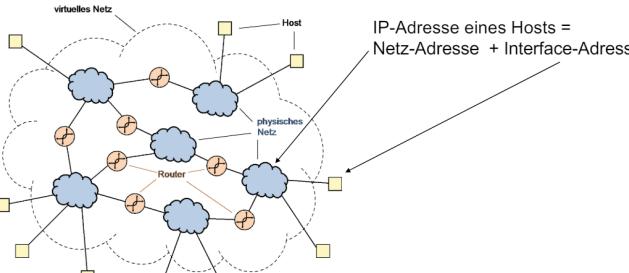
- Router kennt die direkt angeschlossenen Netze seiner Interfaces und genau einen anderen Router, an den er alles schickt, was für andere Netze bestimmt ist
 - Der nächste Router geht genau gleich vor
- Einsatz am „Rand“ von Netzen Hosts, access Router
- Kleine Routing-Tabellen mit jeweils einem Default-Eintrag



Internet Protokolle (IP)

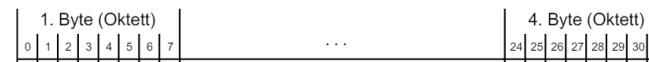
Hierarchische Adressierung

IP-Adressen sind zweistufig hierarchisch

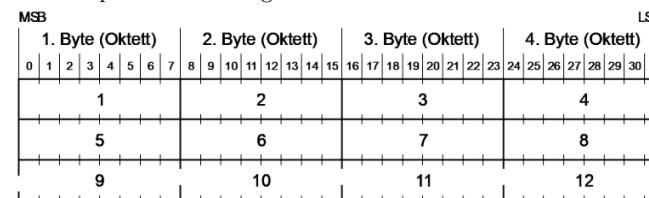


Darstellung von Protokoll Daten

- Aufbau der IP-Adresse

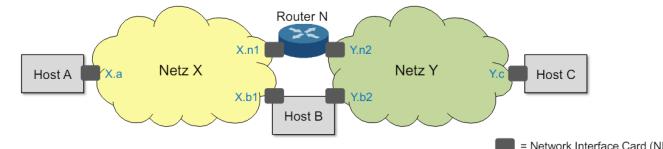


- Prinzip der Darstellung in RFCs



Terminologie

- Sender und Empfänger werden im TCP/IP Referenzmodell als Hosts bezeichnet
- Der Internet Layer stellt ein virtuelles Netz mit einer einheitlichen Adressierung zur Verfügung → IP-Adressen
 - Die IP-Adresse identifiziert ein Host-Interface (und nicht den Host) eindeutig innerhalb eines Netzwerks
 - Jeder Host hat mindestens eine Adresse
 - Multi-Homed Hosts haben mehrere IP-Adressen



Schreibweise

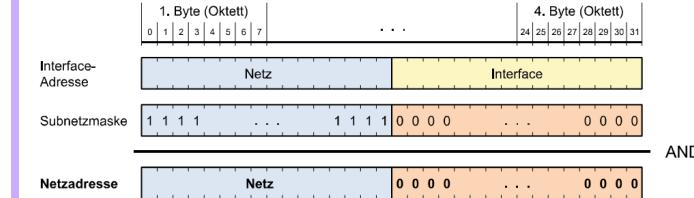
- IP-Adressen werden durch vier Zahlen, unterbrochen durch Punkte, dargestellt
- Jede Zahl entspricht einem Byte in dezimaler Darstellung
- Jede Zahl hat einen Wert im Bereich 0 – 255

1. Byte (Oktett)	2. Byte (Oktett)	3. Byte (Oktett)	4. Byte (Oktett)
0 1 2 3 4 5 6 7	8 9 10 11 12 13 14 15	16 17 18 19 20 21 22 23	24 25 26 27 28 29 30 31
1 0 0 0 0 0 0 0	1 0 0 1 1 0 1 0	0 0 0 1 1 0 1 0	1 1 1 1 0 1 1 0

128 . 154 . 26 . 246

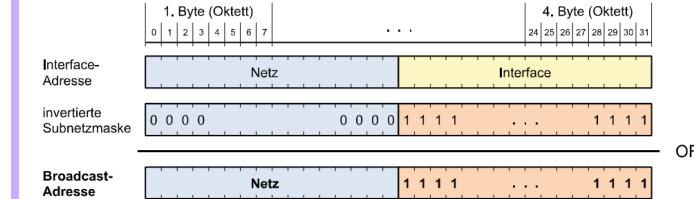
Netzadresse

- Netzadresse ist reserviert: Darf nicht für Interfaces verwendet werden.
- Tiefste Adresse im Subnetz (Interface-Adressbits alle 0)
- Berechnet durch:



Broadcast-Adresse

- Broadcast-Adresse adressiert alle Interfaces in einem Subnetz (reserviert)
- Höchste Adresse im Subnetz (All Ones Broadcast)
- Berechnet durch:



Subnetzmaske bestimmt die Grenze zwischen Netz- und Interface-Adressbits:

1. Byte (Oktett)	2. Byte (Oktett)	3. Byte (Oktett)	4. Byte (Oktett)
0 1 2 3 4 5 6 7	8 9 10 11 12 13 14 15	16 17 18 19 20 21 22 23	24 25 26 27 28 29 30 31
0 0 0 0	0 0 0 0	1 1 1 1	0 0 0 0
0 0 0 0	0 0 0 0	1 1 1 1	1 1 1 1
Netz	Interface	Netz	Interface

Abgekürzte Schreibweise: Anzahl Einsen der Netz-Adresse

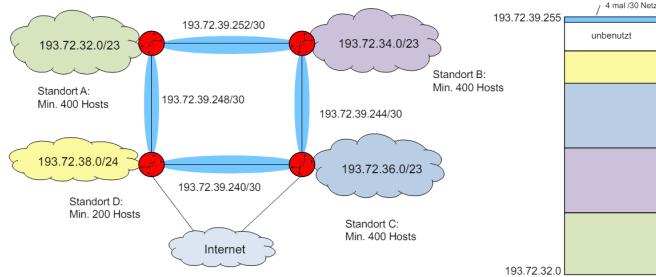
Beispiel:

Netz	Subnetzmaske	Maske
1 0 1 0 0 0 0 0	1 1 1 1 0 0 0 0	1 1 1 1 0 0 0 0
85	16	240
Netz-Adresse		Interface-Adresse
160.85.10.0		160.240.0.0

Abgekürzte Schreibweise für dieses Netz: 160.85.10.0 /20

Flexible Aufteilung eines Netzbereiches

Ein KMU mit 4 Standorten hat von seinem ISP das Netz 193.72.32.0/21 erhalten. Das KMU hat 3 grössere und einen kleineren Standort und will diese redundant verbinden.

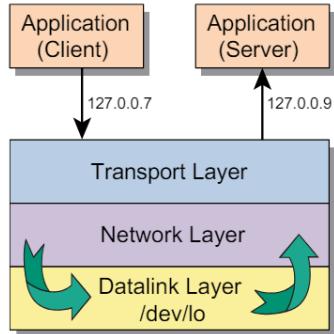


Spezielle IP Adressen

localhost

Loopback-Adressen

- Das gesamte A-Netz 127.0.0.0/8 ist für Loopback-Test reserviert
- Daten werden an ein emuliertes Loopback-Gerät geschickt, das sie direkt zurück gibt (kein Netzwerk-/Interface nötig).



Key Takes Internet/IP

- Die Netzwerkschicht verbindet einzelne, auf Schicht zwei verbundenen Netze, zu einem grossen virtuellen Netzwerk (einem Internet)
 - Interna der einzelnen Netzwerke bleiben für die Endknoten verborgen
 - Die einzelnen Netzwerke können ganz unterschiedliche Technologien verwenden
- Die Netzwerkschicht erfüllt hierbei zwei Hauptaufgaben
 - Routing (Bestimmung der Wege durch das Netz, Aufbau von Routingtabellen)
 - Forwarding (Weiterleitung von Packets gemäss der Routingtabellen)
- Um global Kommunikationsteilnehmer identifizieren (adressieren) zu können, wird ein hierarchisches Adress-Schema verwendet; Routing und Forwarding erfolgt aufgrund der Netzzugehörigkeit von Knoten, nicht aufgrund der Knotenadressen selbst
- Gemäss den Designkriterien bei der Entwicklung der TCP/IP Protokollsuite bietet IP einen verbindungslosen, unzuverlässigen Dienst
 - Erlaubt dafür relativ einfache Implementierung von Routern und bietet ein robustes Verhalten bei Fehlern im Netz (Link-/Komponentenausfall)
- IPV4 Adressen bestehen aus 32 Bit; diese sind in eine Netz- und einen Interface-Teil unterteilt
 - Bitweises AND von Netzmase und IP-Adresse ergibt die Netzadresse
 - Sind alle Bits des Interfaces-Adresssteils gleich „1“, so erhält man die Broadcastadresse im jeweiligen Netz
 - Knoten mit übereinstimmender Netzadresse gehören zum gleichen Netz und können direkt (über Layer 2) miteinander kommunizieren, die Kommunikation zu allen anderen Knoten muss über Router verlaufen
- Routingtabellen definieren, über welche Netzwerk-Interfaces und Router welche Netze erreichbar sind.
 - Sie werden in den Hosts und Routern nach der Grösse der Netzmase abgearbeitet
 - Bei flachem Routing umfasst die Routingtabelle alle bekannten (im Internetwork vorhandenen) Netze
 - Hierarchisches Routing arbeitet mit Default-Einträgen
- Der Default-Eintrag in der Routingtabelle definiert, wohin IP-Pakete geroutet werden sollen, für die keine Eintrag in der Routing Tabelle passt.

IPv4

IP-Header Format

Ein IP-Packet besteht aus einem Header (min. 20 Byte) und Nutzdaten.

- Version** IPv4 / IPv6
- IHL** Header Length in 4-Byte (20 Byte → IHL = 5)
- Type of Service** neu Differentiated Services (DS), Erlaubt Priorisierung, Einteilung der Daten in Verkehrsklassen
 - DSCP: spez. Verhalten bzgl. Weiterleitung
 - ECN: kann drohende Überlast markieren
- Total Length** Länge des IP-Packets (Header + Nutzdaten)
- ID Number** Identifikation des IP-Pakets / Fragmente, erlaubt Identifikation zusammengehöriger Fragmente
- Flags** Kontroll-Flags für Fragmentierung (0/DF/MF)
- Fragment Offset** Gibt an, wo ein Fragment hingehört
- Time to Live** anz. Sek, Hop-Counter, 0 → Paket wird verworfen
- Protocol** Übergeordnetes Protokoll
- Header Checksum** verhindert fehlgeleitete Pakete (nicht Nutzdaten)
- Source Address** Wer das Paket ursprünglich abgesendet hat
- Destination Address** Wer das Paket schliesslich erhalten soll
- Options/Padding** variabel, füllt auf ein Vielfaches von 32Bits auf

1. Byte	2. Byte	3. Byte	4. Byte
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31	Version IHL Type of Service Total Length		
		Identification Number	Flags Fragment Offset
		Time to Live	Protocol IP Header Checksum
			IP Source Address
			IP Destination Address
			Options / Padding

Das unterliegende Netz limitiert die Grösse eines Pakets (Maximum Transfer Unit). Der Sender kennt die MTU der Netze nicht.

Fragmentierung

- Länge der Nutzdaten = Vielfaches von 8 Bytes
- Die Pakete haben die gleiche und grösstmögliche Länge

Reasembly

nutze Flags (0/DF/MF) und Fragment Offset

- Zusammensetzen beim Zielhost
- Letztes Fragment: MF = 0

Feld	Position	Werte	Funktion
	0	0	Reserved, must be Zero
DF	1	0 / 1	May / Don't Fragment
MF	2	0 / 1	Last / More Fragments

Kombination mit DF und MF erlaubt vollständige Rekonstruktion ohne explizite Übertragung der ursprünglichen Paketgrösse

IP-Fragmentierung in heutigen Systemen

Übertragung durch unterliegendes Netz limitiert (maximale Payload)
• Im IP-Kontext als Maximum Transfer Unit (MTU) bezeichnet

- Unterschiedlich für verschiedene Technologien

Fragmentierung in Routern wird vermieden

- Fragmentierung findet im Sender statt
- Entlastet Router von dieser Aufgabe
- Hierzu muss der Sender die kleinste MTU auf dem gesamten Pfad kennen
– Funktionalität siehe ICMP
- Jedes Fragment ist ein vollständiges IP-Paket inklusive Header und wird an den Empfänger weitergeleitet

Das Reassemblieren findet erst im Ziel-Host statt

- Pakete nehmen eventuell unterschiedliche Pfade
- Pakete müssten sonst eventuell später wieder fragmentiert werden

In IPv6 ist keine Fragmentierung "unterwegs" vorgesehen

Kapselung und Adressauflösung

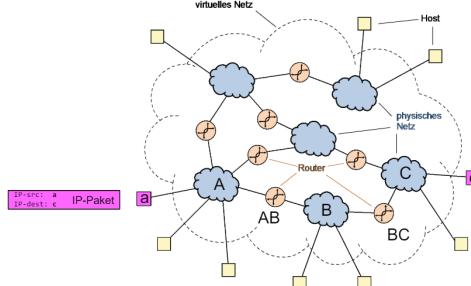
Kapselung eines IP-Pakets im Ethernet Frame

Meist wird heute Ethernet-Encapsulation verwendet

- Das IP-Paket wird direkt im Nutzdatenteil des Frames übertragen
- Das Type Feld des Ethernet Frames erhält den Wert 0800 (hex)
- Die MTU ist damit 1500 Bytes



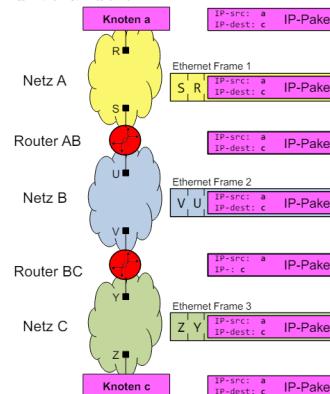
Übertragung eines IP-Pakets mit Encapsulation



Was geschieht bei der Übertragung genau?

- Knoten a sendet ein IP-Paket an Knoten c
– das Paket enthält die IP-Adressen von a und c
- Knoten a konsultiert die Routing Tabelle und sieht:
– dass c über den Router AB erreicht werden kann, und
– Kennt nun die IP-Adresse von Router AB
- Knoten a generiert ein Ethernet Frame, welches an die Hardwareadresse S von Router AB gesendet wird
– a muss aus der IP-Adresse von Router AB die Hardware-Adresse S herausfinden
– **Adressauflösung**
- Router AB empfängt das Ethernet Frame, packt das IP-Paket aus und modifiziert den Header (TTL)
- Router AB konsultiert die Routing Tabelle und sieht:
– dass c über den Router BC erreicht werden kann, und
– Kennt nun die IP-Adresse von Router BC

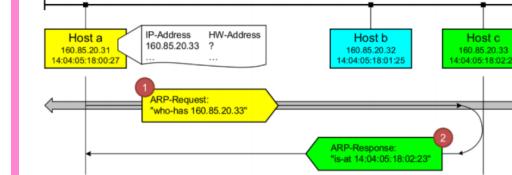
Die IP-Adressen a und c bleiben während der gesamten Übertragung unverändert



Kapselung und Adressauflösung

ARP (Address Resolution Protocol)

- Ermittelt HW-Adresse (MAC) zu einer IP-Adresse

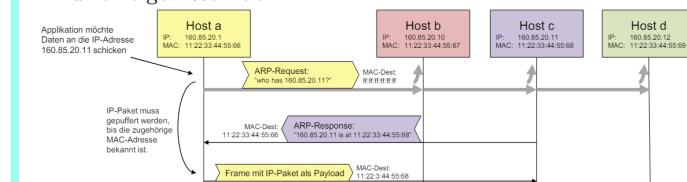


Internet Control Message Protocol (ICMP)

- Übertragungen von Fehlermeldungen oder Informationsaustausch

ARP Grundprinzip von ARP

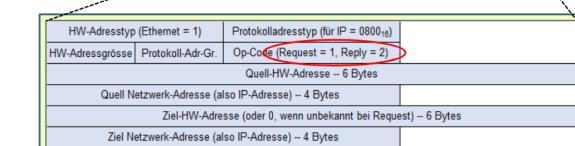
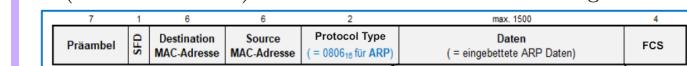
- Für das Senden von Daten an einen durch seine IP-Adresse identifizierten Knoten im lokalen Netz wird dessen Hardwareadresse benötigt.
- Ist diese nicht bekannt, werden alle Knoten im Netz per Broadcast angefragt.
- Der Knoten mit der angefragten IP-Adresse kennt seine eigene Hardwareadresse und sendet sie an den fragenden Knoten zurück
- Die ARP-Tabelle speichert bekannte <IP-MAC> Kombinationen für eine gewisse Zeit



ARP Nachrichtenstruktur

ARP-Request und ARP-Response sind je in genau einem Ethernet Frame enthalten mit Type 0806

- Beim Request ist die Destination Address FF-FF-FF-FF-FF-FF (Broadcast Frame) und die Hardware Address of Target ist 0



ARP Implementierung und Verwendung

- Ein ARP-Request/Response für jedes IP-Paket wäre sehr ineffizient
 - Jeder Knoten führt eine Tabelle (ARP-Cache) mit bekannten HW-Adressen
- Aufgelöste (bekannte) Mappings IP Adresse → Hardwareadresse werden im ARP-Cache für gespeichert
 - Erneuerung nach Ablauf eines Timers, typisch: einige Minuten
- Abfrage/Modifizieren des ARP-Cache mit arp (Windows):
 - arp -a: Anzeigen aller Einträge
 - arp -d ip_addr: Löschen eines Eintrags
 - arp -s ip_addr hw_addr: Setzen eines Eintrags
- Neue / empfohlene Befehle für Linux:
 - ip neigh { add | del | show }

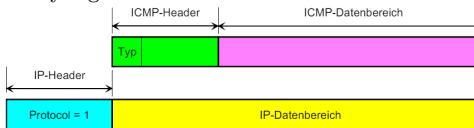
Weitere Verwendung:

- Erkennung von Adresskonflikten
 - Nach einer Adresszuweisung (manuell oder per DHCP) wird ein ARP-Request an die eigene IP-Adresse gerichtet, um zu prüfen, ob kein anderer Host im LAN die Adresse verwendet
 - Falls eine Antwort kommt, liegt ein Adresskonflikt vor
- Erneuerung von Einträgen im ARP-cache
 - Linux Systeme senden in diesem Fall einen ARP-Request als Unicast
 - Reduziert Broadcast-Last im Netz

Internet Control Message Protocol (ICMP)

Übertragung von Fehlermeldungen oder Informationsaustausch auf Internet Layer, z.B.

- Time to live (TTL) hat den Wert 0 erreicht
- Ein Host möchte testen, ob ein anderer Host „up“ ist ICMP Meldungen werden in IP Paketen gekapselt
- Sieht aus wie ein Protokoll eines höheren Layers, welches den Internet Layer verwendet
- ICMP ist aber so eng mit IP verbunden, dass es zum Network Layer gezählt wird

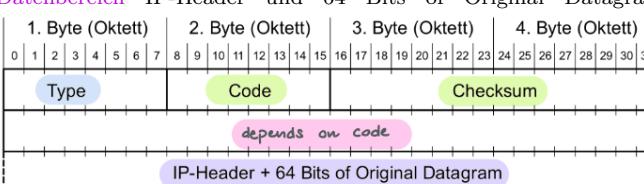


ICMP Format

Header:

- Type ICMP Typ
- Code Message Details
- Checksum Prüfsumme über die ICMP Meldung
- depends on code Unterschiedliche Werte und Verwendung je nach ICMP Typ

Datenbereich IP-Header und 64 Bits of Original Datagram



ICMP Meldungstypen

- ICMP benutzt direkt IP - keine Garantie, dass die Meldungen je ankommen
- Meldungen sind informativ gedacht; Ziel ist nicht eine zuverlässige Übertragung von IP Paketen
- Für eine fehlerfreie Übertragung inklusive Flusssteuerung sind höhere Layer verantwortlich!

ICMP-Typ	Bedeutung (Fehler)
3	Destination Unreachable
5	Redirect
11	Time Exceeded
12	Parameter Problem: Bad IP Header

ICMP-Typ	Bedeutung (Information)
0	Echo Reply
8	Echo
13	Timestamp
14	Timestamp Reply

Destination Unreachable (Fehler)

- IP-Paket kann nicht zum Ziel gebracht werden
- Beispiel: Keine Route zum Ziel-Host vorhanden

Redirect (Optimierung)

- Ein Host H sendet ein IP-Paket an einen ersten Router R1
- R1 stellt fest, dass der nächste Router auf dem Weg zum Ziel R2 ist; R2 ist aber im gleichen Netz wie H und R1 (möglicherweise unvollständige Routingtabelle im Host H)
- R1 sendet an H eine Redirect-Meldung, damit H Pakete fortan direkt an R2 sendet

Time Exceeded (Fehler)

- Router ändert das TTL-Feld im IP-Header von 1 auf 0
- Host hat nicht alle Fragmente erhalten, bevor der Timer abläuft

Parameter Problem: Bad IP Header (Fehler)

- IP Packet Header enthält ungültigen Wert, der nicht verarbeitet werden kann (z.B. nicht existierende IP-Option)
- Echo Request/Reply (Information)

- Host sendet Echo-Request, der adressierte Host antwortet mit Echo-Reply; Reply enthält die gleichen Daten wie Request

Timestamp Request/Reply (Information)

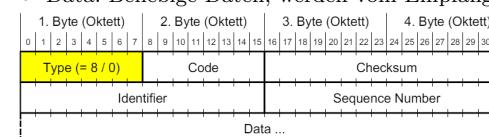
- Wie Echo, aber zusätzlich wird die aktuelle Zeit der Hosts ausgetauscht (32-Bit Wert, Millisekunden seit Mitternacht GMT)

Echo Request/Reply Messages Test, ob Host erreichbar ist

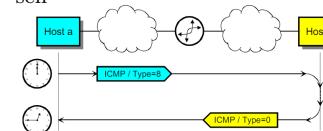
- Host antwortet auf Echo Request (Type 8) mit Echo Reply (Type 0), mit gleichem Inhalt wie der Echo Request

Format

- Identifier: Erlaubt Zuordnung von Reply zu Echo-Request
- Sequence Number: Wird innerhalb eines Identifiers jeweils um 1 erhöht
- Data: Beliebige Daten, werden vom Empfänger gespiegelt



ping verwendet Echo und Echo Reply, um die Erreichbarkeit eines Routers/Hosts zu prüfen; ebenfalls wird die Round-Trip Zeit gemessen



ICMP Destination Unreachable

Vom Router/Zielhost an Absender gesendet, wenn Paket nicht weitergeleitet werden kann

Feld	Wert/Semantik
Type	3
Code	0 = net unreachable, 1 = host unreachable, 2 = protocol unreachable, 3 = port unreachable, 4 = fragmentation needed and DF set, 13 = communication administratively prohibited
Checksum	Prüfsumme über die ICMP Meldung
IP-Header + 64 Bits of Original Datagram	Information für den Empfänger zur Zuordnung der Meldung zu einem gesendeten IP Paket of Original Datagram

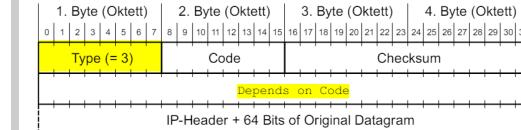
Path MTU Discovery:

Ziel

- Erkennung der kleinsten MTU auf Pfad zwischen Sender und Empfänger (Path-MTU, PMTU)
- RFC 1191 → Path MTU discovery

Zweck

- Vermeidung von Fragmentierung «unterwegs»



Welche Codes werden von einem Router (0,1,4) und welche vom Zielhost (2,3) generiert? Welche vermutlich von einer Firewall (13)?

Path MTU discovery

Annahme, dass die PMTU gleich der lokalen MTU ist

- Senden von IP-Paketen mit Länge=PMTU und mit DF=1
- Empfang von «Destination Unreachable» mit Code 4 «fragmentation needed and DF set»
- PMTU reduzieren auf «Next-Hop MTU»

Die «Next-Hop MTU» erkennt man: Enthalten in Octet 5..8 ("must be zero" stimmt nur, wenn wirklich unused")

ICMP Destination Unreachable

Host 160.85.31.3 versucht, das folgende Paket an Host 160.85.29.99 zu senden (Farben siehe IP-Header def.):

- 4500 0028 8b10 0000 0711 a8a4 a055 1f03 a055 1d63 8b0d 829d 0014 a348 030a 0000 7504 1137 407c 0800

- Erkennen Sie in diesem Paket die IP Adressen von Sender und Destination?
 - Sender : a055 1f03, Destination : a055 1d63

Ein Router kennt keinen Weg und sendet diese Destination Unreachable Message zurück (Farben siehe ICMP-Header def.):

- 4500 0038 8038 0000 fd01 5bc0 a055 821e a055 1f03 0301 4bf7 0000 0000 4500 0028 8b10 0000 0711 a8a4 a055 1f03 a055 1d63 8b0d 829d 0014 a348

- Wie erkennen Sie, dass es sich um eine ICMP Message handelt?
- Protokol: 01
- Wie erkennen Sie den ICMP Typ? Type: 03
- Erkennen Sie die "64 Bytes of Original Datagram"? Original Data

ICMP Time Exceeded

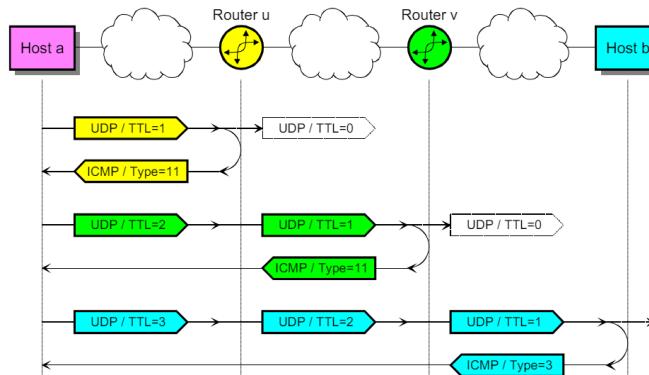
- Type = 11
- unused (must be 0)

Wird in diesen 2 Fällen gesendet:

- Router setzt TTL-Feld von 1 auf 0
 - Paket wird verworfen und der Absender informiert (Code = 0)
- Zielhost kann ein fragmentiertes Paket nicht innerhalb nützlicher Zeit reassemblieren
 - Fragmente werden verworfen und der Absender informiert (Code = 1)

traceroute erlaubt, den Weg zu einem beliebigen Host (oder einem fehlerhaft konfigurierten Router auf diesem Weg) zu finden

- UDP Datagramme an hohe Destination Portnummer (zufällig gewählt, default 33434)
- Erstes Datagramm: TTL := 1
 - Erster Router setzt TTL auf 0, verwirft Paket und sendet Time Exceeded Message zurück
 - Erste Router ist bekannt
- Nächstes Datagramm: TTL := 2
 - Zweiter Router ist bekannt etc...
- ...
- Zielhost kann Zielpot nicht erreichen
 - Destination Unreachable Message (Code = 1) an Absender
 - Zielhost ist erreicht
- Um die Entfernung zu den einzelnen Routern/Zielhosts zu bestimmen, wird zugleich noch die Round-Trip Zeit gemessen



IPv6

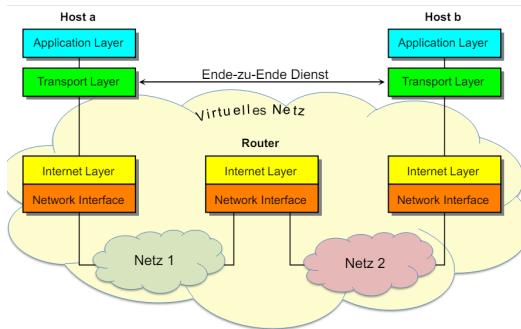
- IPv6 ist in RFC 2460 spezifiziert
- 128-bit Adressen; diese werden mit je zwei Bytes in Hex-Darstellung notiert und durch Doppelpunkte getrennt
- IPv6 verwendet Extension Headers, um den Basic Header zu vereinfachen
- Ein Interface kann mehr als eine IPv6 Adresse haben
 - Ein Interface hat in der Regel eine lokale und zwei globale IPv6 Adressen:
 - Eine MAC-basierte und eine nicht von der Hardware abhängige.
- IPv6 verwendet zur Abfrage der Layer-2 Adressen NDP statt ARP
- Domain Name Service (DNS)
 - IPv4 stellt an den Resolver Anfragen nach A-Records
 - IPv6 stellt an den Resolver Anfragen nach AAAA-Records
- Link-Local Adressen sind nur im lokalen Layer-2 Segment gültig; sie werden nicht geroutet
- hat sich nicht durchgesetzt weil:
 - nicht so einfach lesbar wie IPv4
 - Viele Probleme von IPv4 konnten gelöst werden
 - IPv6 ist nicht rückwärtskompatibel, daher teuer zu implementieren, und benötigt viel zusätzliches Fachwissen von Netzwerkadministratoren

Key Takes

- Der IP-Header besteht aus 20 Bytes (ohne Optionen)
- Um über Netze mit verschiedenen Maximum Transfer Units (MTU) arbeiten zu können, unterstützt IP Fragmentierung und Reassembly
 - Heute wird in der Regel beim Sender fragmentiert und im Ziel-Host reassembliert
 - Path MTU discovery mittels ICMP kann verwendet werden, um die kleinste MTU auf dem Weg zum Ziel-Host zu identifizieren
- IP-Pakete werden in Ethernet Frames gekapselt und von jedem Router wieder ausgepackt und erneut gekapselt.
 - Dazu muss der Router die Layer-2 Adresse (MAC-Adresse) des nächsten Routers/Hosts kennen (ARP-Cache) oder erfragen (ARP-Request)
- ICMP wird verwendet, um Fehler innerhalb der Netzwerkschicht zu behandeln (keine Retransmissions)
 - ICMP-Nachrichten werden in IP-Pakete gekapselt, werden aber dennoch der Netzwerkschicht zugeordnet

Transport Layer

Schicht 4: Transportschicht



Transportlayer

Der Transport Layer bildet auch die Schnittstelle zwischen dem Betriebssystem (Kernel Space) und den Anwendungen (User Space). Der Zugriff auf die Funktionen des Transport Layers erfolgt via einer klar definierten Schnittstelle (Sockets).

Kapselung

- Die Applikationsdaten werden von den Protokollen des Transport Layers in ein IP-Paket gekapselt
- Das gekapselte Protokoll wird im IP-Header im Feld Protocol angegeben
- Das "Protocol"-Feld unterscheidet UDP und TCP Daten

Adressierung der Applikation durch Port Nummern

Der Client adressiert mit der Destination Port Nummer die gewünschten Server-Applikation

- sonst weiß das TCP/UDP-Modul im Empfänger nicht, welche Applikation gemeint ist
- für die Source Port Nummer verwendet der Client (meist) eine zufällige Port Nummer im Bereich >1'023 (wird vom Betriebssystem vergeben)

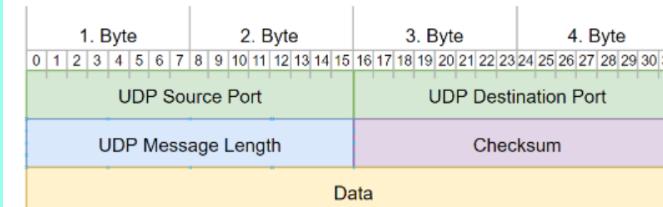
UDP - User Datagram Protocol

UDP dient dem Multi- und Demultiplexen der Datagramme zu den Applikationen.

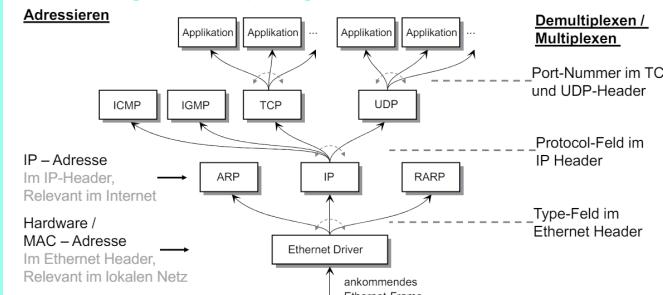
- Verbindungslos
- Unzuverlässig

UDP-Header

- Source Port Sendende Applikation
- Destination Port Applikation des Empfängers
- Message Length Länge des Datagramms
- Checksum Prüfsumme über einen Pseudo-Header, UDP-Header und Daten
 - kann Null sein
- Pseudo-Header: IP Source- und Destination Address, Protocol Feld, Länge des Datagramms
 - so können fehlgeleitete Datagramme erkannt werden
 - z.B. aufgrund eines Bit-Flip



Adressierung und Multiplexing



Port-Nummern

- System Ports (Well-Known)** Feste Port-Nummern, für bekannte Appl. reserviert
- User Ports (Registered)** Reservierter Bereich für herstellerspezifische Appl.
- Dynamic / Private Ports** Frei verfügbare Ports

System Ports	User Ports	Dynamic Ports
0 - 1023	1024 - 49'151	49'152 - 65'535

TCP - Transmission Control Protocol

TCP Eigenschaften

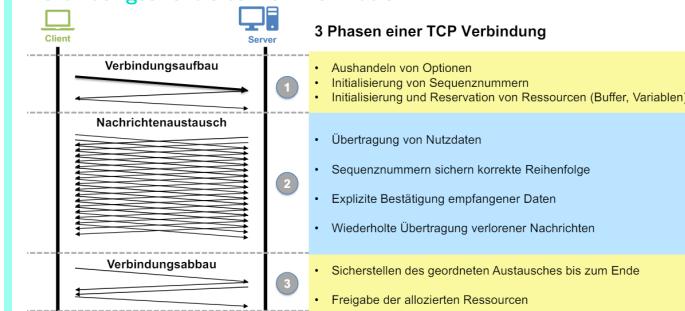
- Verbindungsorientierte Übertragung: Zuerst wird eine Verbindung zwischen Client- und Serveranwendung aufgebaut
- Zuverlässiger Verbindungsauflauf: Bevor eine TCP-Verbindung steht, muss dies von beiden Endpunkten aktiv bestätigt werden
- Hohe Zuverlässigkeit: Die Daten kommen ohne Datenverlust und in der richtigen Reihenfolge auf der anderen Seite an
- Vollduplexübertragung: Gleichzeitige, voneinander unabhängige, Übertragung in beiden Richtungen möglich
- Stream-Schnittstelle: Die Anwendung sendet/empfängt eine unstrukturierte Byte-Folge
- Graceful Termination (Verbindungsabbau): TCP gewährt die Zustellung aller Daten auch beim Verbindungsabbau
- Punkt-zu-Punkt Kommunikation: Zwei Applikationen tauschen Daten aus. Konzepte wie Multicast oder Broadcast existieren nicht.

Zuverlässigkeit

TCP muss diverse Probleme lösen:

- Eine Verbindung soll zuverlässig auf- und abgebaut werden können
 - Verbindungsauflauf, Verbindungsabbau
- TCP-Nachrichten können verloren, verfälscht, dupliziert und deren Reihenfolge vertauscht werden – trotzdem sollen alle Daten korrekt, vollständig und in der richtigen Reihenfolge auf der anderen Seite an die Applikation weitergegeben werden
 - Sequenznummern, Adaptiver Timeout, Sliding Window Protokoll
- Der Empfänger soll nicht mit Daten überschwemmt werden, d.h. der Sender soll sich an die Möglichkeiten des Empfängers anpassen
 - Flow Control mit Advertized Window Size
- Das Netz dazwischen soll nicht überlastet werden, damit ein vernünftiger Durchsatz für alle möglich ist
 - Congestion Control mit Slow Start Algorithmus

Verbindungsorientierte Kommunikation



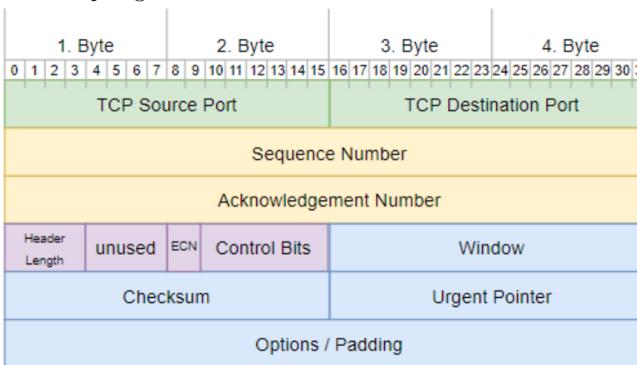
TCP Header

Header Specs

- TCP Header Länge: Mindestens 20 Bytes plus mögliche Optionen (+ max. 40 Bytes)
- Eine TCP-Verbindung besteht aus je einem Datenstrom in jeder Richtung
- Header beinhaltet Information für die "Vorwärtsrichtung"(Sequenznummer etc.) und für die "Rückwärtsrichtung"(Acknowledgement Nummer, Window)

TCP-Header Format

- Sequence-Nr. Nummer zur Ordnung der Segmente
- Acknowledgement-Nr. $n + 1 \rightarrow$ Daten bis und mit n korrekt und vollständig angekommen
- Data Offset Gibt an wo Daten beginnen / enden
- ECN-Flags Explicit Congestion Notification
 - Bit 8: CWR (Congestion Window Reduced)
 - Bit 9: ECE (ECN-Echo)
- Control Bits Verbindungsaufl- und -abbau (Bits 10-15) URG: Urgent Pointer ACK: Acknowledgement Number PSH: Push (sofort ohne buffern weiterleiten) RST: Reset (Verbindung zurücksetzen oder geschlossenen Port signalisieren) SYN: Synchronize (Verbindung aufbauen) FIN: Verbindung abbauen
- Window Verfügbarer Puffergrösse (so viele Bytes dürfen noch gesendet werden)
- Urgent Pointer URG = 1 \rightarrow Position der wichtigen Daten
- Options Häufigste Verwendung: MSS (Maximum Segment Size) die empfangen werden kann



TCP Verkehrssteuerung

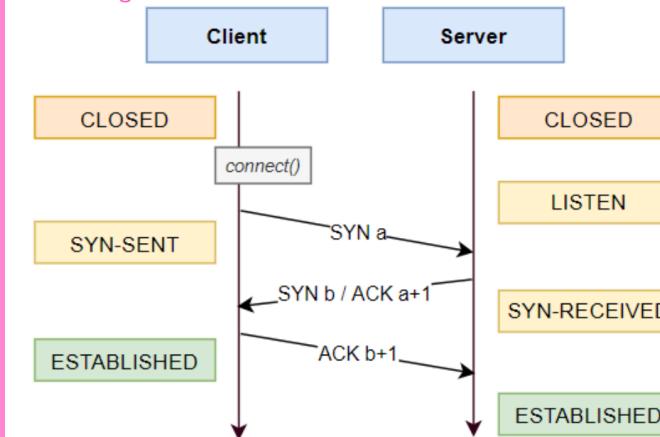
Nachrichtenaustausch

- Unabhängig für jede Richtung:
- Sequence Numbers identifizieren jedes Byte des gesendeten Nutzdatenstroms
 - Sicherstellen der richtigen Reihenfolge der Daten
 - Erkennen verloren gegangener Daten
 - Acknowledge Numbers identifizieren korrekt empfangene Bytes (der Gegenrichtung)
 - Bestätigung korrekt empfangener Daten
 - Erkennen verloren gegangener Daten
 - Flags steuern Verbindungsaufl- und -abbau, signalisieren Gültigkeit von Informationen im Header und besondere Situationen.
 - SYN/FIN: Verbindungsaufl- und -abbau
 - ACK: Acknowledge Number im empfangenen Segment ist gültig
 - PSH: Daten sollen schnellstmöglich an Applikation weitergegeben werden

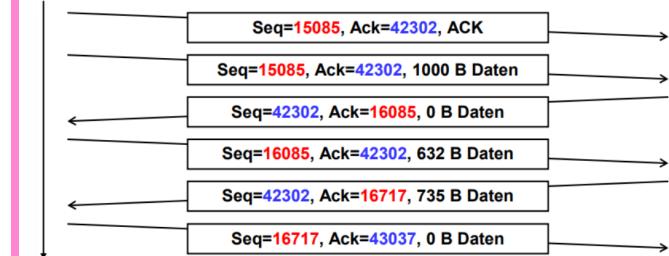
Zustände

- LISTEN Auf Anforderung warten
- SYN-SENT Anforderung geschickt
- SYN-RECEIVED Anforderung erhalten
- ESTABLISHED Verbindung besteht
- FIN-WAIT-1 Abbauanforderung geschickt
- FIN-WAIT-2 Abbauanforderung bestätigt
- CLOSE-WAIT Auf Lokale Verbindung warten
- LAST-ACK Verbindungsabbau bestätigt
- TIME-WAIT Letzte Bestätigung gesendet

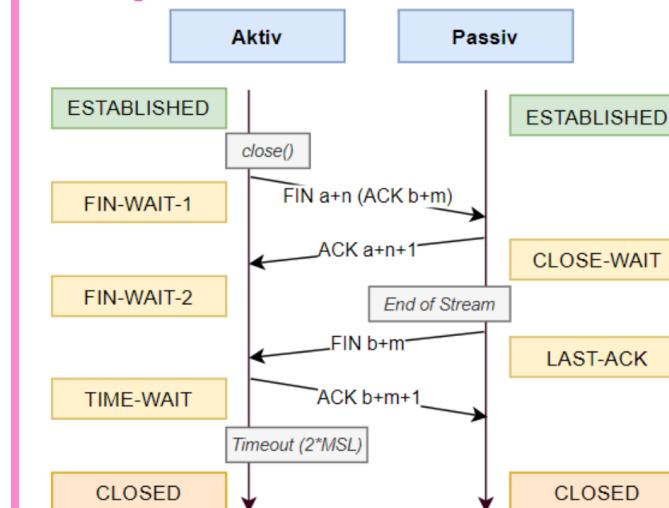
Verbindungsauftbau



Datenaustausch

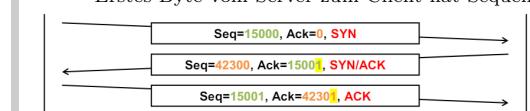


Verbindungsabbau



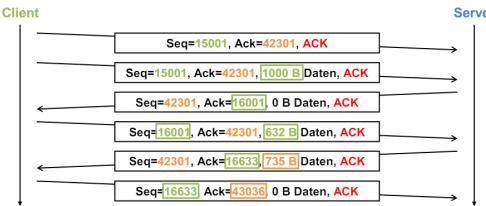
Verbindungsauftbau:

- Server „horcht“ (LISTEN) auf einer bestimmten Port Nummer (z.B. 80 für einen HTTP Server)
- Client sendet Segment mit SYN=1 und zufälliger initialer Sequenznummer a (z. Bsp. 15'000) (ACK=0, weil Acknowledgement Nummer ungültig)
- Server bestätigt Sequenznummer mit Acknowledgement Nummer a+1 (15'001) und ACK=1 und wählt zufällige initiale Sequenznummer b (z. Bsp. 42'300) und setzt SYN=1
- Client bestätigt b mit Acknowledgement Nummer b+1 (42'301)
 - Erstes Byte vom Client zum Server hat Sequenznummer a+1
 - Erstes Byte vom Server zum Client hat Sequenznummer b+1



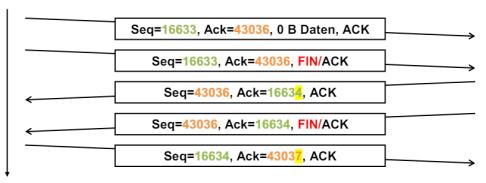
Während des Datenaustausches werden TCP-Nachrichten bi-direktional ausgetauscht

- Sequenznummer: Position des ersten Bytes der Daten im gesamten TCP-Datenstrom
- Acknowledgement Nummer: Sequenznummer des nächsten erwarteten Bytes
- ACK Flag: immer gesetzt

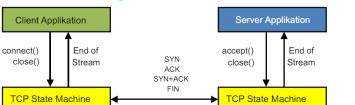


Beide Seiten können den Verbindungsabbau einleiten

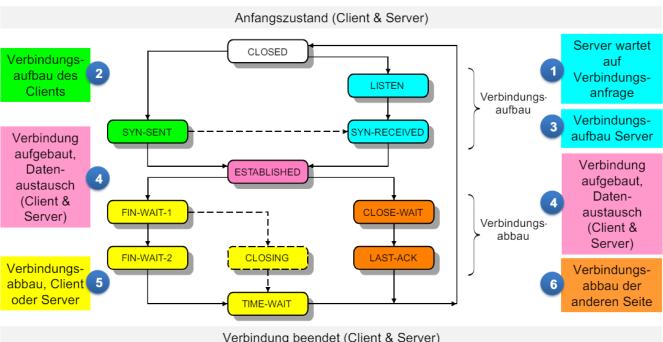
- Ist eine Richtung geschlossen (FIN, ACK), so können in die andere Richtung immer noch Daten gesendet werden; dieser Verbindungsstatus wird als Half-Closed bezeichnet.
 - In Richtung der "geschlossenen" Verbindung wird nicht mehr kommuniziert (Acknowledge number mismatch)
- Falls die zweite Seite die Verbindung auch schliesst, können die 3. und die 4. Nachricht zusammengefasst werden → FIN/ACK



Zustandsdiagramm

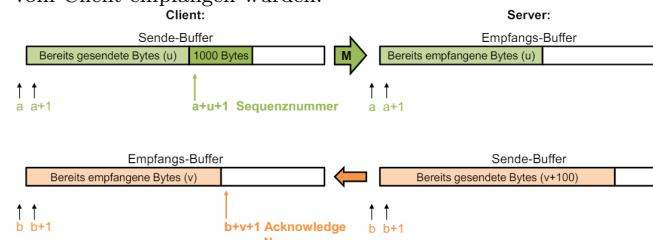


- Um die verschiedenen Zustände zu durchlaufen, existiert beim Client und Server pro Verbindung je eine TCP State Machine
- Applikationen beeinflussen die TCP State Machines mit Systembefehlen
 - listen(), connect(), close()
- Die TCP State Machines signalisieren sich Events mit speziellen Flags der Control Bits im TCP-Header einer TCP-Nachricht:
 - SYN, ACK, FIN



Datenaustausch

Geben Sie die Seq- und Ack-Nummern der Meldung M (1000 Bytes von Client zum Server) an und zeichnen Sie die entsprechenden Positionen ein. Beachten Sie, dass 1000 Bytes vom Server noch nicht vom Client empfangen wurden.

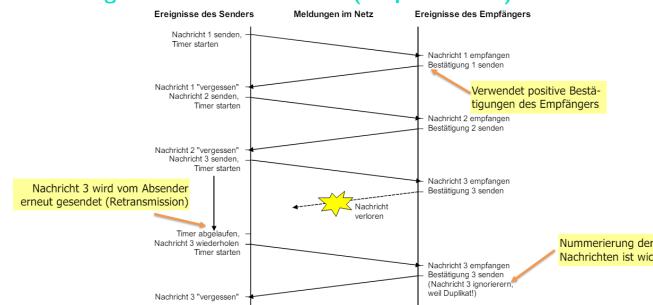


TCP Adaptive Elemente

Umgang mit dynamischen Situationen

- Erkennung von verlorenen Telegrammen (Round Trip Time)
 - Kann je nach Auslastung des Netzes stark variieren
 - Wie soll der Timeout für die Detektion von Nachrichtenverlust gewählt werden?
- Überlast des Empfängers (Fluss-Steuerung, Flow Control)
 - Der Empfänger erhält u.U. von Hunderten oder Tausenden von Clients Daten geschickt.
 - Wie kann der Empfänger den Sender steuern?
- Überlast des Netzes (Überlast-Steuerung, Congestion Control)
 - Selbst wenn der Empfänger ausreichend Performance und Puffer zur Verfügung hat, kann das Netz durch "Querverkehr" überlastet werden.
 - Wie kann der Sender das erkennen und das Netz schützen?

Erkennung verlorener Nachrichten (Stop and Wait)



Round Trip Time Erkennung von verloren gegangenen Telegrammen
Um Fehler Paketverluste und andere Fehler zu verhindern, werden Pakete nach einer bestimmten Zeit erneut übertragen, wenn keine Bestätigung gesendet wurde. Um diese Zeit zu optimieren, misst TCP bei jeder aktiven Verbindung die Round-Trip Time (RTT).

- Gewichteter Mittelwert SRTT (Smoothed Round-Trip Time)
 $\alpha = 0.125 : SRTT_n = (1 - \alpha) \cdot SRTT_{n-1} + \alpha \cdot RTT_n$
- Streuung RTTVar des SRTT der Abweichungen
 $\beta = 0.25 : RTTVar_n = (1 - \beta) \cdot RTTVar_{n-1} + \beta \cdot SRTT_n - RTT_n$
- Retransmission Time-Out RTO
 $RTO_n = SRTT_n + 4 \cdot RTTVar_n$

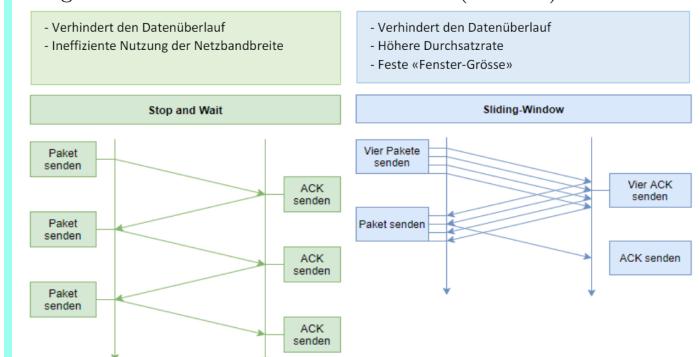
Fluss-Steuerung

Überlast des Empfängers

Problem: Der Sender sendet Daten schneller, als diese vom Empfänger verarbeitet werden können. Folgen:

- Überlauf im Empfangsbuffer
- Daten im Empfänger gehen verloren

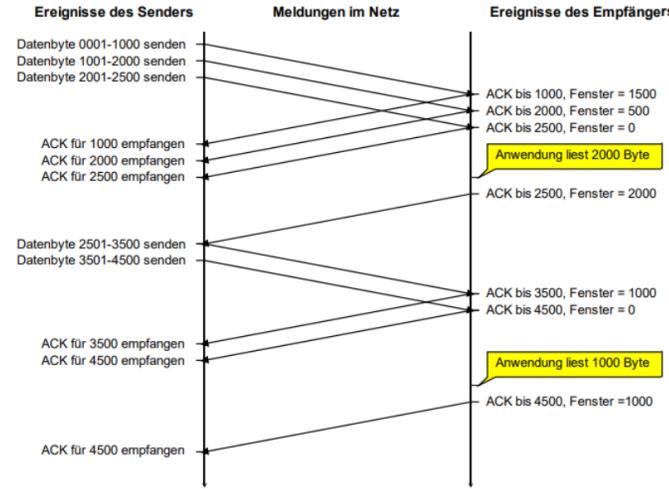
TCP verwendet den Sliding-Window Mechanismus um Kanalauslastung zu erhöhen. Beide Seiten einen Buffer (Window).



Sliding-Window TCP

- Beide Richtungen arbeiten unabhängig voneinander
- Fenstergrösse wird in Anzahl Bytes angegeben
- Verbindungsauftakt: Initiale Fenstergrösse wird der anderen Seite mitgeteilt
 - Typische Werte: 16 / 32 / 64 KB
- Pufferplatz im Empfänger wird alloziert (der Empfänger wählt selbst die Fenstergrösse!)
- Mit jedem ACK wird der verfügbare Pufferplatz (in Bytes) mitgeteilt und damit die Fenstergrösse dynamisch angepasst
- Erhält ein Sender ein Fenstergrösse von 0 Bytes mitgeteilt, dürfen keine weiteren Daten gesendet werden
- Ist im Empfangsbuffer wieder Pufferplatz vorhanden, wird erneut eine Bestätigung mit diesem Pufferplatz an die andere Seite gesendet (= aktuelle Fenstergrösse)

Fluss-Steuerung bei TCP

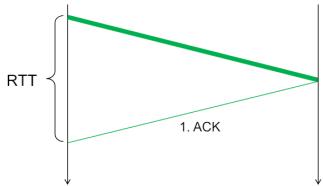


Bandwidth Delay Product (TCP-Puffergrößen)

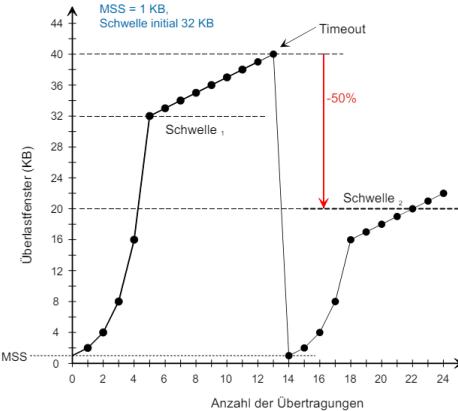
Wie gross sollten die Sende- und Empfangsbuffer gewählt werden, um eine TCP-Verbindung nicht auszubremsen?

$$BDP(\text{bits}) = RTT(\text{sec}) \cdot \text{Bandbreite}(\text{bps})$$

RTT = Round-Trip-Time



Congestion Control



Überlast des Netzes

TCP benutzt den Paketverlust als Masseinheit für Überlastung und reagiert durch Absenken der Übertragungsrate (Slow Start). Dadurch kann die Überlastung überwacht und verhindert werden. Hierfür pflegt jeder Sender zwei Fenster (vom Sender gewährtes Fenster, Überlastungsfenster). Das Minimum der Fenster stellt die Anzahl Bytes dar, die gesendet werden können.

Wichtig: Der Sender kombiniert das Congestion Window mit den Informationen zur Flow Control vom Empfänger und schickt unbestätigte Daten bis zum Erreichen von: min {Congestion Window, Advertised Window}

Zusammenfassung Transport Layer

Vergleich Layer 4 und Layer 2

Herausforderungen zur Zuverlässigkeit zwischen Ethernet (Schicht 2) und TCP (Schicht 4):

- Spalte 1: Potentielle Probleme / Fehlersituationen
- Spalte 2/3: Charakteristika derselben auf Schicht 2 und Schicht 4
- Spalte 4: Massnahmen bei TCP, um diese Probleme zu lösen

Problem	Schicht 2	Schicht 4	Massnahmen bei TCP
Nachrichtenverlust	$P_{\text{Verlust}} = \text{FER}$	$P_{\text{Verlust}} > \text{FER}$	Positives ACK
Telegramm-Reihenfolge	fix	kann variieren	Sequenznummern
Round Trip Time	konstant, $\mu\text{s} \dots \text{ms}$	variabel, $\text{ms} \dots \text{s}$	Adaptiver Retransmission Timeout
Überlast des Empfängers	kommt vor	kommt vor	Sliding Window mit dynamischer Fenstergrösse
Überlast des Netzwerks	direkt beobachtbar (Medium)	nur indirekt beobachtbar	Slow Start (Congestion Window)
Neustart von Hosts	direkt beobachtbar	nur indirekt beobachtbar	3 Weg Handshake, Initialisierung Sequenznr.

Key Takes Transport Layer

- Die Transportschicht bietet Applikationen einen Ende-zu-Ende Dienst
 - Port Nummern bei TCP / UDP werden verwendet um Applikationen zu identifizieren
 - Server Applikationen warten (=listen) an bekannten (well-known) Ports
 - Der Vorgang, lokal Daten an verschiedene Instanzen höherer Protokollschichten zu verteilen, wird als Multiplexing / Demultiplexing bezeichnet und findet auf allen Schichten statt
 - * „Type“ bei Ethernet, „Protocol“ bei IP, „Port“ bei TCP/UDP
- UDP gibt die Eigenschaften von IP (verbindungslos, unzuverlässig) praktisch unverändert weiter
 - Zusatzfunktionalität: Multiplexen / Demultiplexen aufgrund der Port Nummer und Checksumme über die gesamten Daten (IP schützt nur den Header)
- TCP bietet einen verbindungsorientierten, zuverlässigen Dienst
 - Für die Applikation sehr ähnlich zum Lesen / Schreiben eines Files
 - Verbindung von einer Client-Application zu einer Server-Applikation
 - Verbindungsauftbau wird immer vom Client initialisiert
 - Nach dem Verbindungsauftbau ist die Kommunikation symmetrisch, beide Kommunikationspartner können das Schliessen der Verbindung initialisieren
- Ein UDP Datagramm / TCP Segment wird in genau ein IP Paket eingefügt
- TCP verwendet folgende Massnahmen, um die sichere Kommunikation zu gewährleisten:
 - 32-Bit Sequenznummern – jedes Byte im Datenstrom ist eindeutig identifiziert
 - 3-Wege Handshake beim Verbindungsauftbau mit zufälliger Initialisierung der Sequenznummern
 - Kontrollierter Verbindungsabbau mit der Möglichkeit, ausstehende Daten zu senden
 - Adaptive Timeouts basierend auf Wert und Varianz der gemessenen Round-Trip Zeiten
 - Schutz des Empfängers vor Überlast durch dynamische Anpassung der Fenstergrösse beim Sliding Window Protokoll (Advertised Window, wird dem Sender vom Empfänger mitgeteilt)
 - Schutz des Netzwerks vor Überlast durch Congestion Control (im Sender) mit dem Slow Start Algorithmus

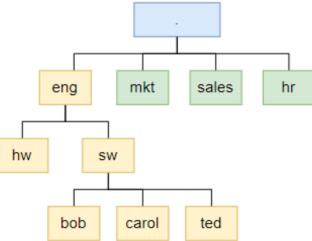
Application Layer

Netzwerk-Applikationen und Protokolle

Domain Name System

DNS - Domain Name Space

- Leserliche Darstellung von IP-Adressen (Name Resolution)
- Hauptdomäne = Root
- Der Fully Qualified Domain Name (FQDN) muss eindeutig sein, Beispiel: sw.eng.
- Geschwisterknoten dürfen nicht den gleichen Namen haben

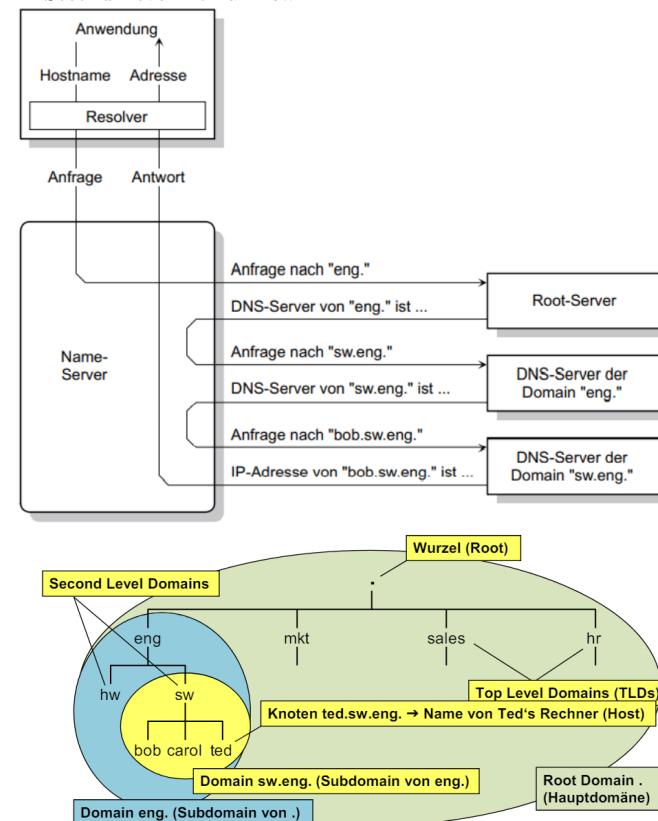


Verwaltung von Domains

- Das DNS wird verteilt betrieben (verteilt, nicht repliziert): Name Server
- Ein Name Server ist meist für eine Zone verantwortlich
 - Zone: separater administrierter Subtree des DNS, oft eine Domain
 - Ein Name Server kennt
 - * die IP-Adressen zu den Hostnamen in seiner Zone
 - * die IP-Adressen der Name Server seiner Subdomänen, falls diese nicht in seiner Zone liegen
 - * die IP-Adressen von Root und TLD Name Server, um beliebige Abfragen zu erlauben
- Aus Redundanzgründen mindestens zwei Name Server für eine Zone
 - Primary (Master) Name Server und Secondary (Slave) Name Server
- Ein NS kann eine Unterzone seiner Zone weiter delegieren

DNS-Abfragen auswerten

- DNS verwendet Port 53 (UDP)
 - Resolver: lokale Software, die mit dem Name Server kommuniziert
- Beispiel: Anwendung benötigt die IP-Adresse von bob.sw.eng.
- FQDN: bob.sw.eng.
 - Root: .
 - Top Level Domain: eng
 - Second Level Domain: sw



DNS Record Types

Der "Record Type" enthält Information, welche Daten angefragt beziehungsweise in einer Antwort vom Name Server mitgeteilt werden

Type	Beschreibung / Funktion	Definiert in
A	IPv4 Adresse des gesuchten Hosts (32 Bit)	RFC 1035
AAAA	IPv6 Adresse des gesuchten Hosts (128 Bit)	RFC 3596
MX	Mail Exchange (Mail Server)	RFC 1035 / 7505
NS	Name Server (Name Server Name für eine Zone)	RFC 1035
CNAME	Canonical Name (primärer Name) für einen Alias zum Host	RFC 1035
TXT	Text Record, in Antworten für verschiedene Angaben verwendet	RFC 1035

Reverse DNS Lookup

Authentisierung: Ein Server identifiziert/authentifiziert einen Client anhand des Namens, nicht anhand der IP-Adresse

DHCP - Dynamic Host Configuration Protocol

Bezug IP-Adresse

Wie erhält ein Knoten seine IP-Adresse?

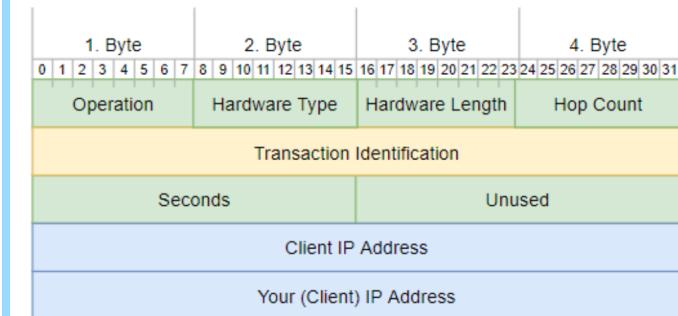
- Lokal konfiguriert (static IP)
- Bezug der IP-Adresse über das Netzwerk
 - DHCP – erlaubt dynamische Zuteilung aus dem lokalen Adressbereich

Dynamische Zuweisung von IP-Adressen

- Client verlangt eine IP-Adresse (DHCP Request)
- DHCP-Server erteilt eine freie Adresse für definierte Lease Time, oft 10 Minuten (DHCP-Response)
- Vor Ablauf der Lease Time muss der Lease (vom Client) erneuert werden
- Client, der das Netz verlässt, wird Lease nicht erneuern → Adresse wieder frei

DHCP - Dynamic Host Configuration Protocol

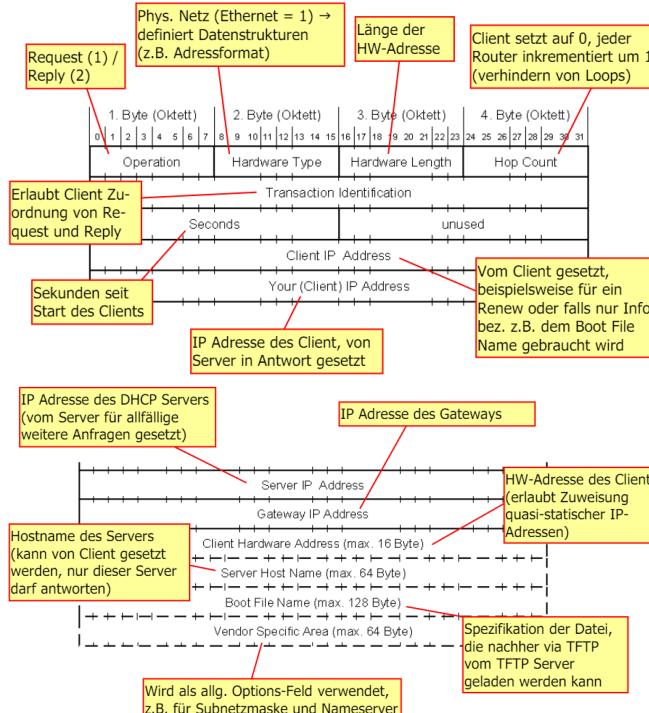
- Dynamische Zuweisung von IP-Adressen
- Reserviert nur IP's von aktiven Geräte



Ablauf DHCP

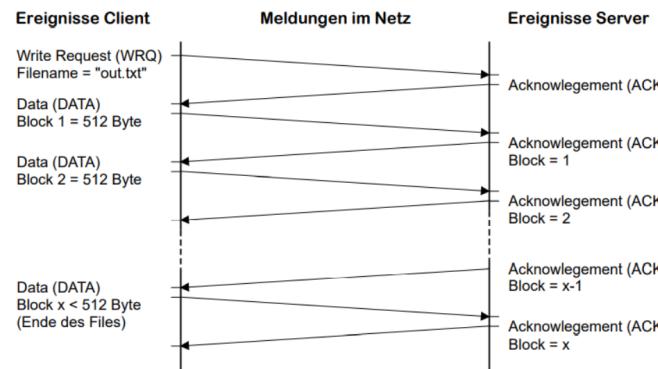
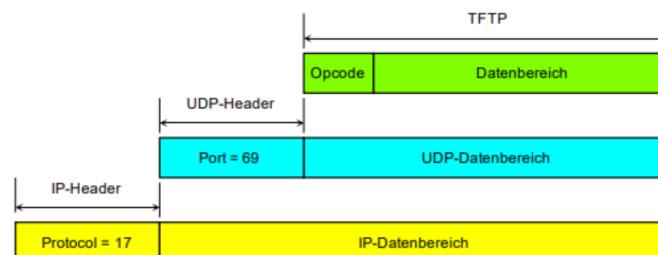
1. Client sucht DHCP Server mittels Broadcast
2. DHCP Server antwortet (DHCP offer)
3. Der Client wählt einen Server und fordert eine Auswahl der angebotenen Parameter (DHCP request)
4. Der Server bestätigt mit einer Message, welche die endgültigen Parameter enthält
5. Vor Ablauf der Lease-Time erneuert der Client die Adresse.

DHCP Paketformat



TFTP - Trivial File Transfer Protocol

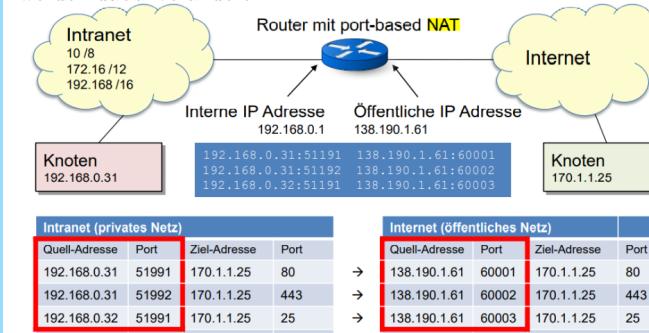
Basiert auf UDP



NAT - Network Address Translation

NAT (Port Mapping)

- NAT (Historisch) Sicherheit durch «Verstecken» von lokalen Adressen
- NAPT (Port Trans.) Lokale IP-Adresse → Öffentliche IP-Adresse
NAT verletzt das Konzept der OSI-Layer, da eine Network-Funktion auf den Transport-Header zugreift. IP-Adresse und Portnummer werden dabei verändert.



NAT ist das letzte Prüfungsrelevante Thema!!!

Key Takes Applikationsprotokolle

- Das Domain Name System (DNS) erlaubt übersetzt Hostnamen in IP Adressen und umgekehrt
 - Besteht aus einem hierarchischen DNS Name Space
 - Das DNS wird auf einer grossen Anzahl Name Server verteilt betrieben, ein Name Server ist jeweils für eine Zone verantwortlich (z.B. zhaw.ch)
- DHCP erlaubt einem Rechner, seine IP Konfiguration von einem Server zu beziehen
- TFTP ist ein einfaches, aber zuverlässiges File Transfer Protocol, welches z.B. diskless Systemen dazu dient, das Betriebssystem Image vom Server zu beziehen
- HTTP erlaubt den Zugriff auf verteilte Dokumente, die mittels Uniform Resource Locator (URL) eindeutig adressiert werden
- Network Address Translation (NAT) erlaubt die Wiederverwendung privater IP-Adressen

HTTP - Hypertext Transfer Protocol

- WWW basiert auf HTTP

Funktionsweise von HTTP

- Basiert auf TCP, Port 80
- ASCII-Basiert, MIME-Typen, Codierungen
- Transaktionsbasiert: HTTP Request → HTTP Response