Rechnernetze & Verteilte Systeme

Ludwig-Maximilians-Universität München

Sommersemester 2017

Prof. Dr. D. Kranzlmüller

MNM TEAM

Kapitel 4: Vermittlungsschicht

ISO/OSI-Schicht 3

(Engl. Network Layer)

MNM TEAM

Inhalt von Kapitel 4

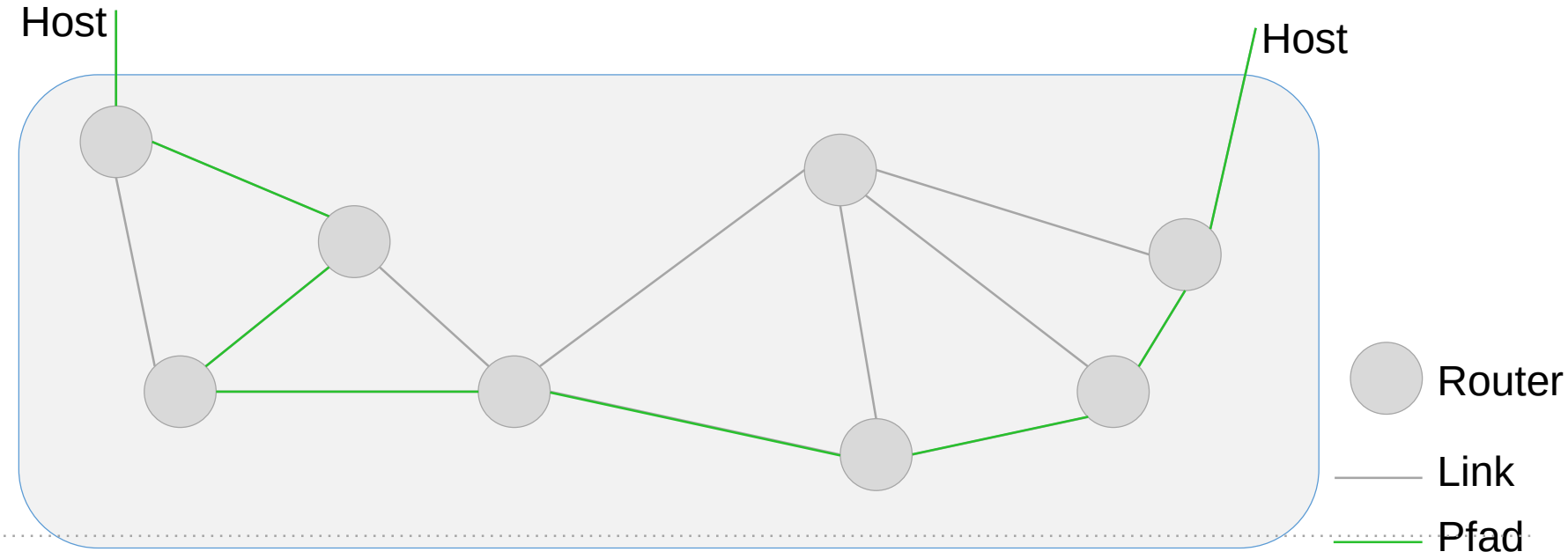
1. Ziele und Rahmenbedingungen
2. Wegewahl auf anderen Schichten
3. Vermittlung (Pfadschaltung)
4. Wegewahl (Routing)
5. Internet Protokoll (v4)
6. Internet Protokoll (v6)

Einordnung von Kapitel 4

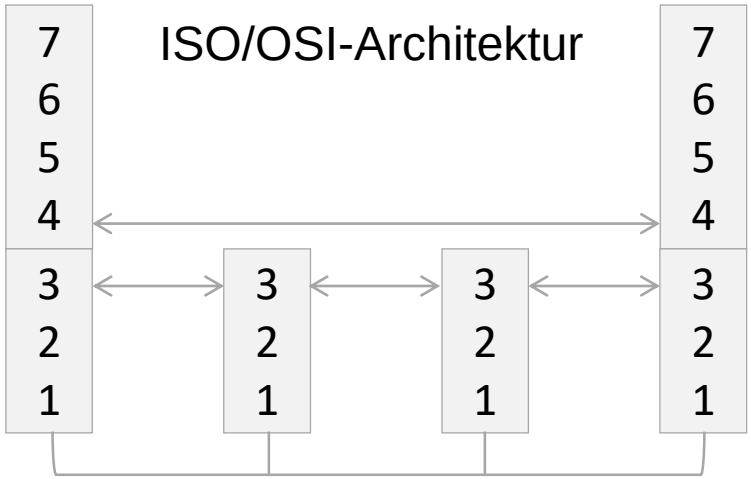
- Im Kapitel 3 haben wir uns mit der Transportschicht (ISO/OSI-Schicht 4) beschäftigt: Netzunabhängiger Transport von Nachrichten zwischen Endsystemen
- Kapitel 4: Vermittlungsschicht: Übertragung von Paketen von der Quelle zum Ziel
- Vermittlung (Pfadschaltung) und Wegewahl (Routing) als Dienst
- Internet Protokoll (IP): v4 und v6

Ziele und Rahmenbedingungen der Vermittlungsschicht

Wegewahl, Dienstgüte, Verbindungsart, Sicherungsschicht als Rahmenbedingung



Hauptziel: Pfadschaltung und Wegewahl



Dienst: Verbindungslos oder verbindungsorientiert? (1/2)

- Prinzipiell: Auch auf der Vermittlungsschicht verbindungslos oder verbindungsorientiert
- Bei einem verbindungsorientierten Vermittlungsprotokoll müssen sich die Netzknoten den Pfad, den Daten einer gegebenen Verbindung nehmen sollen, merken.

  ➔Alle Pakete der Verbindung werden auf demselben Pfad übertragen.
- Beispiel für verbindungsorientierte Vermittlung: Telefonnetz

Dienst: Verbindungslos oder verbindungsorientiert? (2/2)

- Im Internet: Verbindungsloses Internet Protokoll
  - IP-Pakete werden unabhängig voneinander befördert.
  - Zwischenknoten müssen sich keine Zustandsinformationen merken.
- Vorteil: Beim Ausfall einzelner Netzknoten werden Pakete einfach umgeleitet (anstatt alle Verbindungen, die über den Knoten laufen, neu aufzubauen).

Verbindungsart und Ende-zu-Ende Argument

- Entscheidung für verbindungslosen Ansatz im Internet als Folge der Ende-zu-Ende Argumentation
- Wenn Verbindungen als inhärent unzuverlässig angesehen werden und verbindungsorientierte Aufgaben bereits auf der Transportschicht übernommen (TCP) werden, gibt es kaum Vorteile dies auf der Vermittlungsschicht erneut zu implementieren
- Nachteil: Realisierung von Dienstgüte schwierig (z.B.: für Echtzeitdatenverkehr wie Sprache und Video)
  - ➔ „Unter der Oberfläche“ entwickelt das Internet durchaus verbindungsorientierte Eigenschaften (Bsp.: VLAN – Schicht 2)

Schicht 3 Dienstgüte (Engl. Quality of Service – QoS)

- Dienstgüte auf Vermittlungsschicht gewinnt an Bedeutung: Streaming-Dienste und IP-Telefonie
- Schicht 3 Dienstgüteparameter:
  - Verbindungsaufbauwahrscheinlichkeit (Blockierwahrscheinlichkeit)
  - Verbindungsaufbauzeit
  - Durchsatz der Schicht-3-Verbindung
  - Nachrichtenübertragungszeit
  - Schwankung in der Übertragungszeit (engl.: Jitter)
  - Restfehlerrate

Die Sicherungsschicht als Rahmenbedingung (1/2)

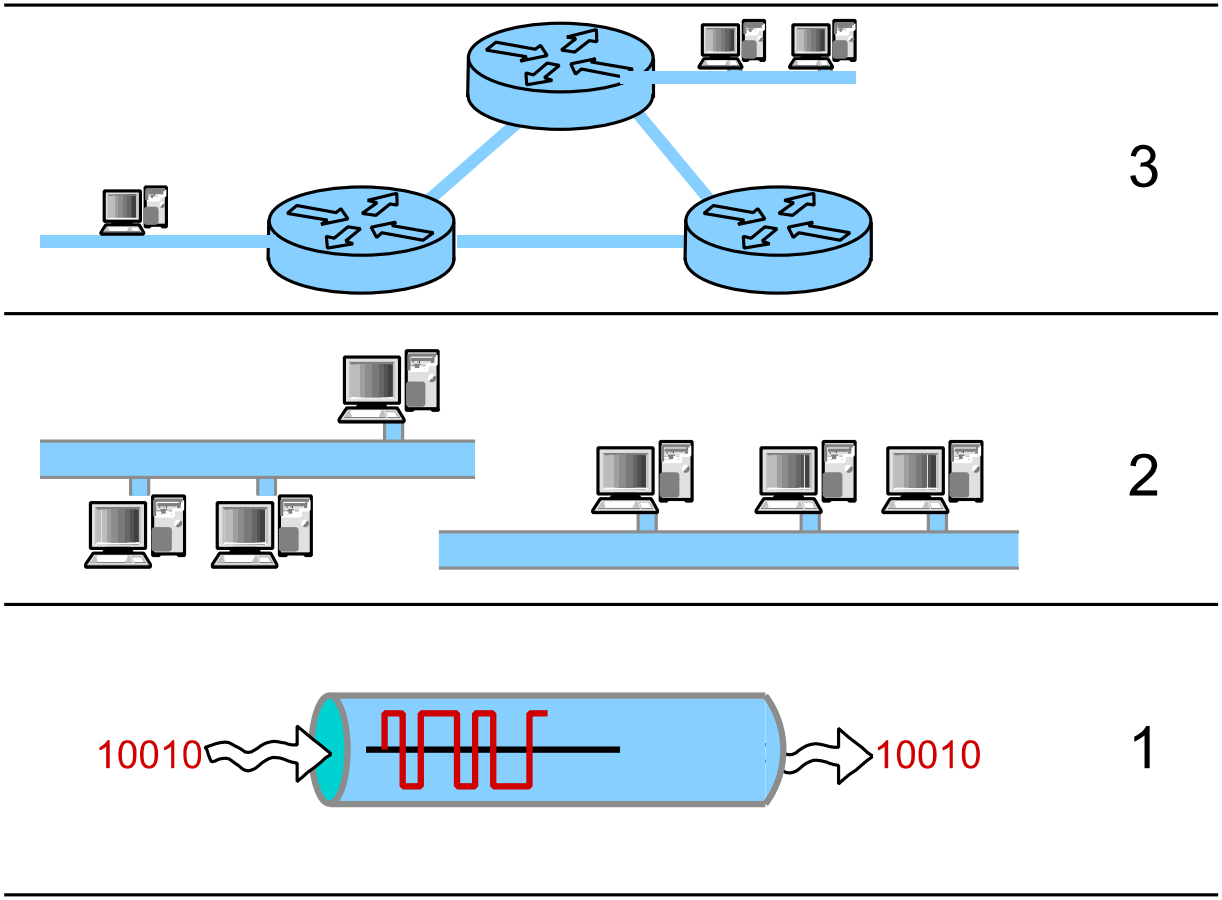
- Vermittlungsschicht bietet ihre Dienste *unter Verwendung der Dienste der nächst tieferliegenden Schicht* an: ISO/OSI-Schicht 2: Sicherungsschicht
- Verweis auf Kapitel 5 (Hardwarenahe Schichten):

  Übertragung über ein einzelne Wegabschnitte (von einem Knoten über ein Medium zum nächsten Knoten) als *gesicherten Data Link/Logical Link.*

Die Sicherungsschicht als Rahmenbedingung (2/2)

- Übertragung über einzelne Wegabschnitte (von einem Knoten über ein Medium zum nächsten Knoten) als gesicherten Data Link/Logical Link.
  - Medien- und Übertragungstechnik unabhängig
  - Bits zu Blöcken/Rahmen (engl. Frames) zusammengefasst.
  - Fehlererkennung (und ggf. Korrektur) der Rahmenübertragung
- Vermittlungsschicht setzt also voraus, dass Nachrichtenpakete (bis zu einer gewissen Größe – Stichwort MTU) fehlerfrei an Nachbarknoten weitergeleitet werden können.

Netzbildung



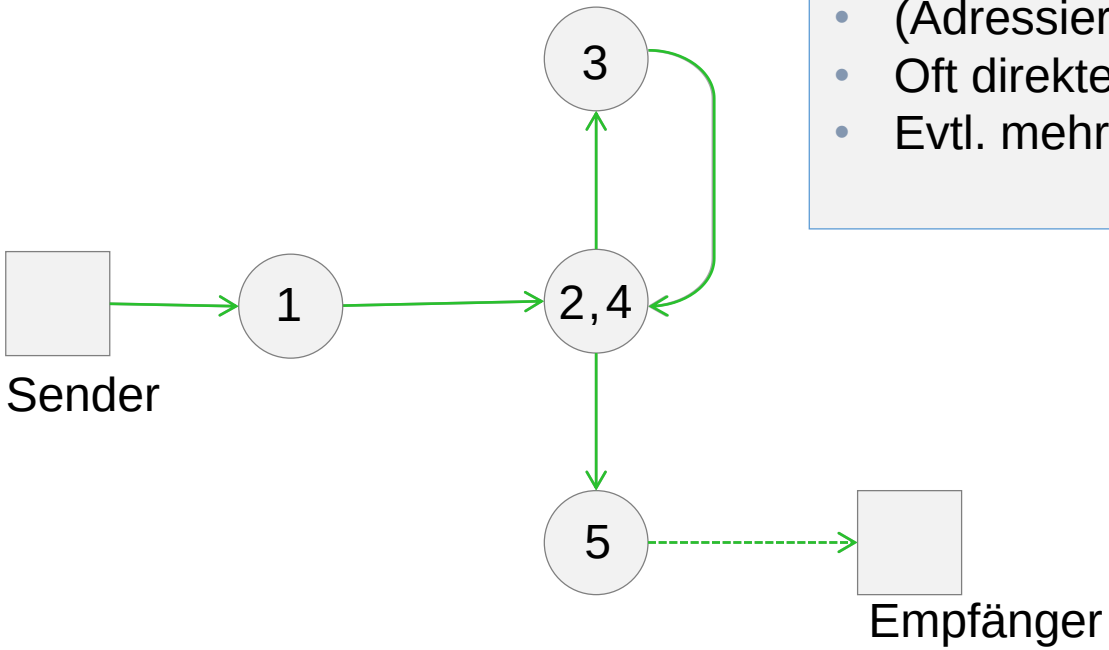
Kapitel 4.2
Wegewahl auf anderen Schichten

E-Mail, Bridges und Switches, IP-Router

Einordnung/Motivation

- Wegewahl und Vermittlung können (technologieabhängig) auch auf anderen Schichten vorkommen (schichtunabhängiges Konzept).
- Überblick anhand von drei kleinen Beispielen:
  - ISO/OSI-Schicht 7: E-Mail-Relays
  - ISO/OSI-Schicht 3: IP-Router
  - ISO/OSI-Schicht 2: Bridges und Switches

Beispiel: Anwendungsschicht: E-Mail-Relay (ISO/OSI-Schicht 7)



Beispiel: Vermittlung von E-Mails

- Wegewahl anhand von DNS
- (Adressierung: Domain Name)
- Oft direkte Zustellung
- Evtl. mehrere Hops

E-Mail-Relays in der Praxis

Received: from mailrelay1.lrz-muenchen.de (mailrelay1.lrz-muenchen.de [129.187.254.106])
by **mail.nm.ifi.lmu.de** (8.12.11/8.12.11/Linux MNM 0.1) with ESMTP id l9CAC5v7004532
for <danciu@nm.ifi.lmu.de>; Fri, 12 Oct 2007 16:16:05 +0200
Received: from lxmhs06.lrz-muenchen.de (lxmhs06.lrz-muenchen.de [10.156.6.203]) by **mailrelay1.lrz-muenchen.de** with ESMTP for danciu@nm.ifi.lmu.de; Fri, 12 Oct 2007 16:16:04 +0200
Received: from mailrelay1.lrz-muenchen.de ([10.156.6.201])
by **lxmhs06.lrz-muenchen.de** (lxmhs06.lrz-muenchen.de [10.156.6.203]) (amavisd-new, port 10024)
with ESMTP id mDqRYsUc8VAp for <danciu@nm.ifi.lmu.de>;
Fri, 12 Oct 2007 16:16:03 +0200 (CEST)
Received: from mail.gmx.net (mail.gmx.net [213.165.64.20]) by **mailrelay1.lrz-muenchen.de** for danciu@nm.ifi.lmu.de; Fri, 12 Oct 2007 16:16:02 +0200
Received: (qmail invoked by alias); 12 Oct 2007 10:11:01 -0000
Received: from yyyyyyyyyy.dip.t-dialin.net (EHLO ASMCore2Duo) [217.238.48.17x]
by **mail.gmx.net** (mp042) with SMTP; 12 Oct 2007 16:15:01 +0200
From: "NN" <xxxxxx@gmx.de>
To: "'Vitalian A. Danciu'" <danciu@nm.ifi.lmu.de>
Subject: Frage Rechnernetze 1

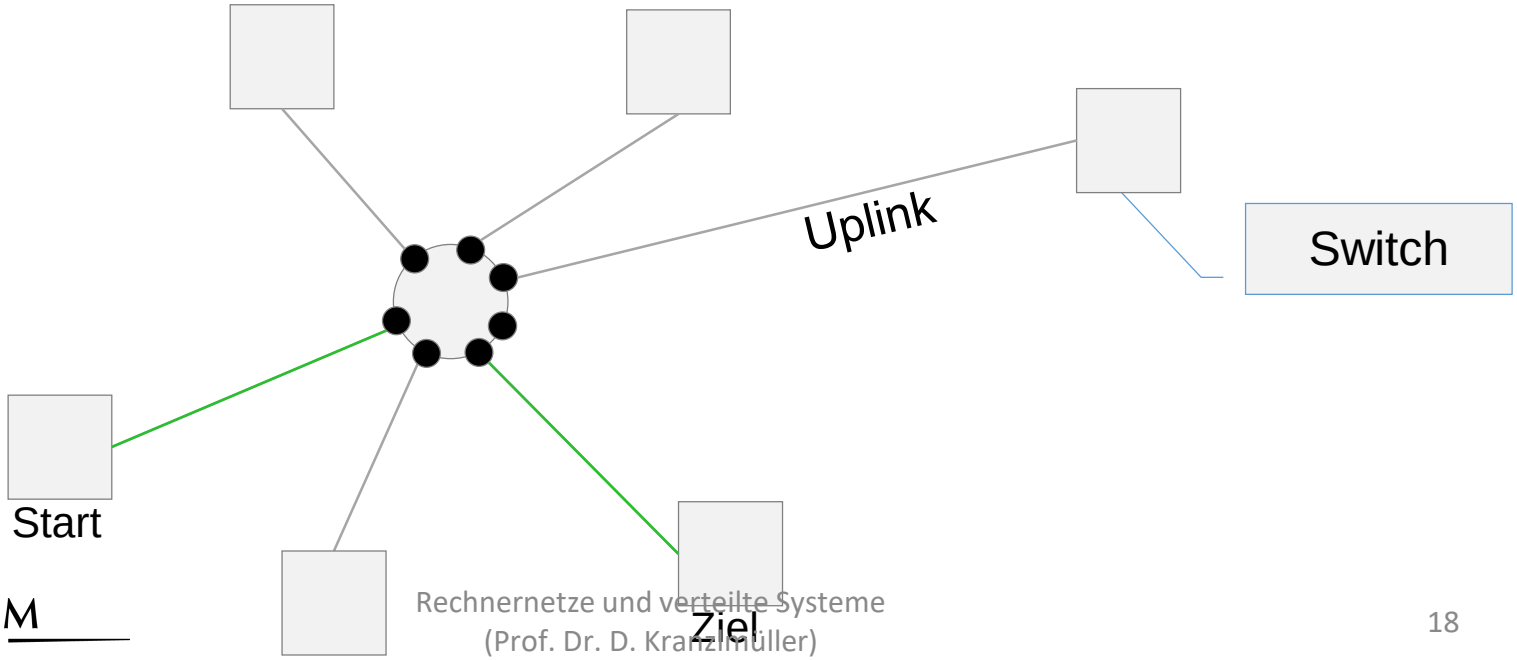
TEXT

Achtung!
Verwechslungsgefahr
Ports auf Schicht 4

Beispiel: Sicherungsschicht: Switch (ISO/OSI-Schicht 2)

Beispiel: Wegewahl in einem Switch (Schicht 2)

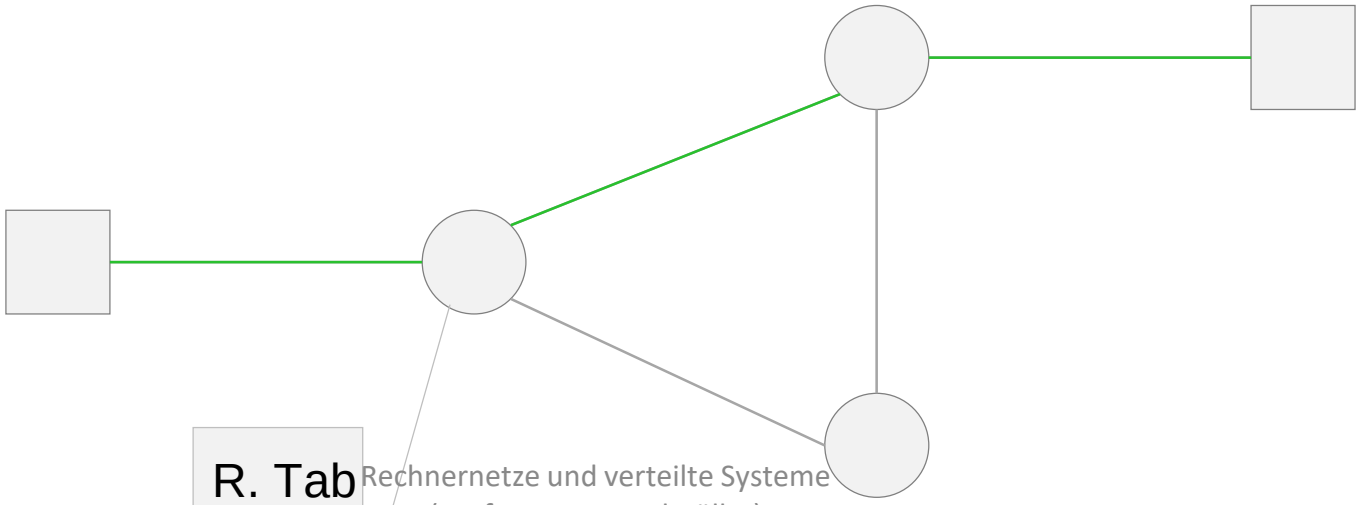
- Ports: Anschlussstellen (Ein-/Ausgänge) für ein Medium
- Switch findet Zuordnung (Adressierung: MAC-Adresse)
  - anhand von Broadcast
  - anhand von bereits beobachtetem Netzverkehr
- Switch erlernt Zuordnung



Beispiel: Vermittlungsschicht: IP-Pakete (ISO/OSI-Schicht 3)

Beispiel: Wegewahl bei IP-Paketen (Schicht 3)

- Router trifft Entscheidung zur Weiterleitung von Paketen
- Weg bestimmt durch (Adressierung: IP-Adresse)
  - Regeln in Routingtabellen
  - Angaben des Senders (selten)
- Routingprotokolle: zum Aushandeln optimierter Wege
- Nach technischen Gesichtspunkten
- Nach organisatorischen/finanziellen Gesichtspunkten



R. Tab

Kapitel 4.3: Vermittlung (Pfadschaltung)

Leitungsvermittlung, Nachrichtenvermittlung, Paketvermittlung

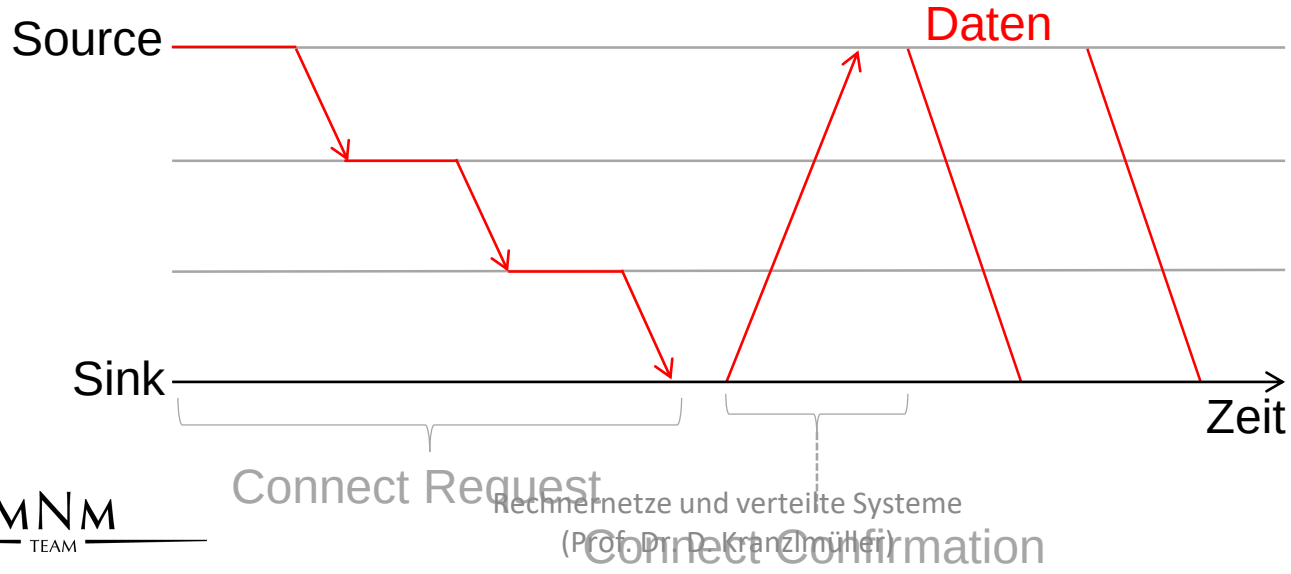
Einordnung/Motivation

- Wegewahl (Routing): Finde anhand von (netzglobalen) Adressen einen Pfad durch das Netz
- Vermittlung (Pfadschaltung): Übertragung der Daten über die einzelnen Wegstücke für einen Ende-zu-Ende Pfad (Beispiele: als eine große Nachricht oder aufgeteilt in viele kleine Pakete).
- Vermittlungsverfahren:
  - Leitungsvermittlung
  - Nachrichtenvermittlung
  - Paketvermittlung

Leitungsvermittlung (Engl. Circuit Switching)

Leitungsvermittlung, Durchschaltung, engl. circuit switching

- Vor Übertragung: Aufbau des Pfades von Sender zu Empfänger
- Dedizierter Kanal für die gesamte Datenphase
- typisch ist gleicher Grenzdurchsatz für alle Links
- typisch gekoppelt mit verbindungsorientiertem Dienst
- Beispiel: *POTS, ISDN-B-Kanäle, Telefonnetz (früher)*



Vor- und Nachteile der Leitungsvermittlung

Nachteile

- Ende-zu-Ende Pfad muss vollständig eingerichtet sein, bevor Daten übertragen werden können
- Bricht eine Vermittlungseinrichtung zusammen, gehen auch alle Verbindungen, die darüber laufen, verloren.
- Wenn Kanal vollständig vergeben ist, kann er von sonst niemandem mehr verwendet werden (Besetztzeichen).

Vorteile

- Ist der Pfad gefunden, besteht ein dedizierter Ende-zu-Ende Kanal.

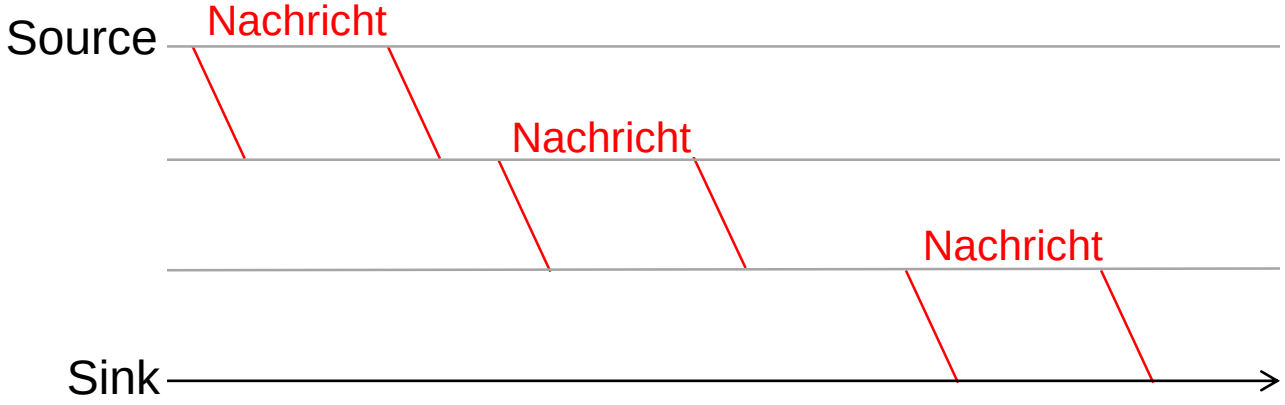
➔Es gibt keinerlei Warteschlangen in den Knoten.

➔Es ist einfach eine geforderte Dienstgüte zu garantieren.

Nachrichtenvermittlung (Engl. Message Switching)

Wird auch als „Store and forward“ Verfahren bezeichnet

- Die ganze Nachricht wird zum jeweils nächsten Knoten geschickt und dort zwischengespeichert, bis dieser sie wieder weiterleitet.
- Streckenstücke sind im Allgemeinen nicht homogen.
- Schwankende Verarbeitungs- und Warteschlangenverzögerungen.
- Beispiel: *E-Mail*



Vor- und Nachteile der Nachrichtenvermittlung

Nachteile

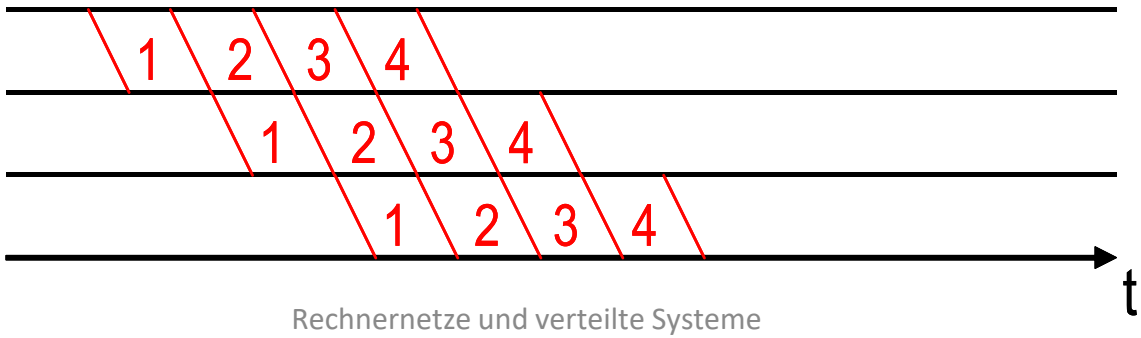
- Übertragung auf dem jeweils nächstem Wegstück beginnt erst, wenn die Nachricht vollständig eingetroffen ist
- Lange Nachrichten können einzelne Wegstücke (für andere Nutzer) lange blockieren.
- Notwendigkeit großer Zwischenspeicher in den Knoten.

Vorteile

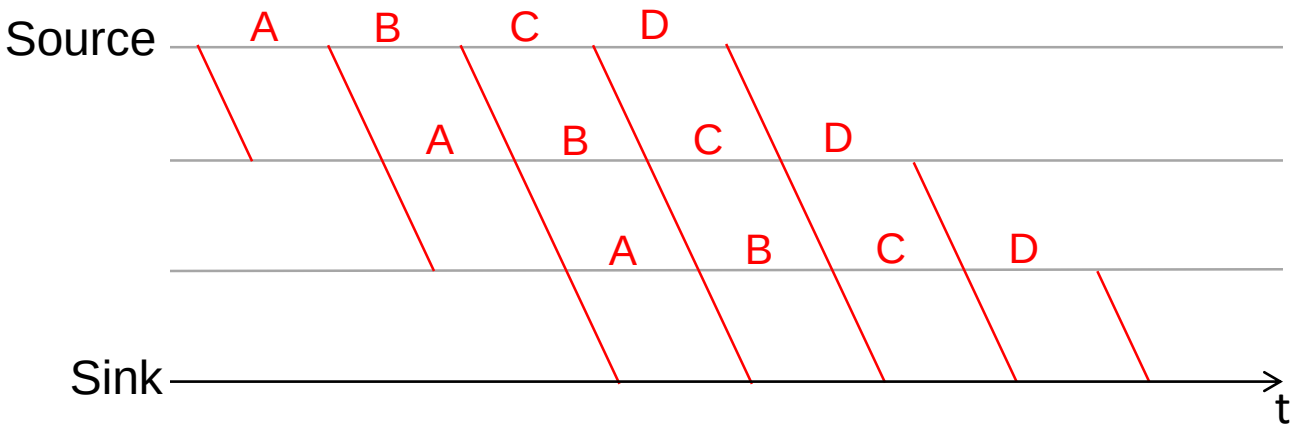
- Nachrichtenvermittlung vor allem dann sinnvoll, wenn die Konnektivität zwischen den Vermittlungseinrichtungen nicht dauerhaft gegeben ist verzögerungstolerante Netze
- Bsp.: Low-Earth-Orbit Satelliten, U-Boote

Paketvermittlung (Engl. Packet Switching)

- Zerlegung der Nachrichten in etwa gleichlange Pakete.
- Senden der Nachrichten **paketweise** nach dem *Store and Forward* Verfahren (vorherige Folie).
- Bessere Leitungsauslastung und geringere Latenz (bzgl. Eintreffen des 1. Pakets) durch *Pipelining Effekt*
- Reihenfolgeproblem
- Verbindungsorientiert sowie verbindungslos möglich
- Beispiel: Internet , ATM



Paketvermittlung (Animation)



Source muss noch senden: A, B, C, D

Unterwegs:

Sink hat erhalten: A, B, C, D

Vor- und Nachteile der Paketvermittlung

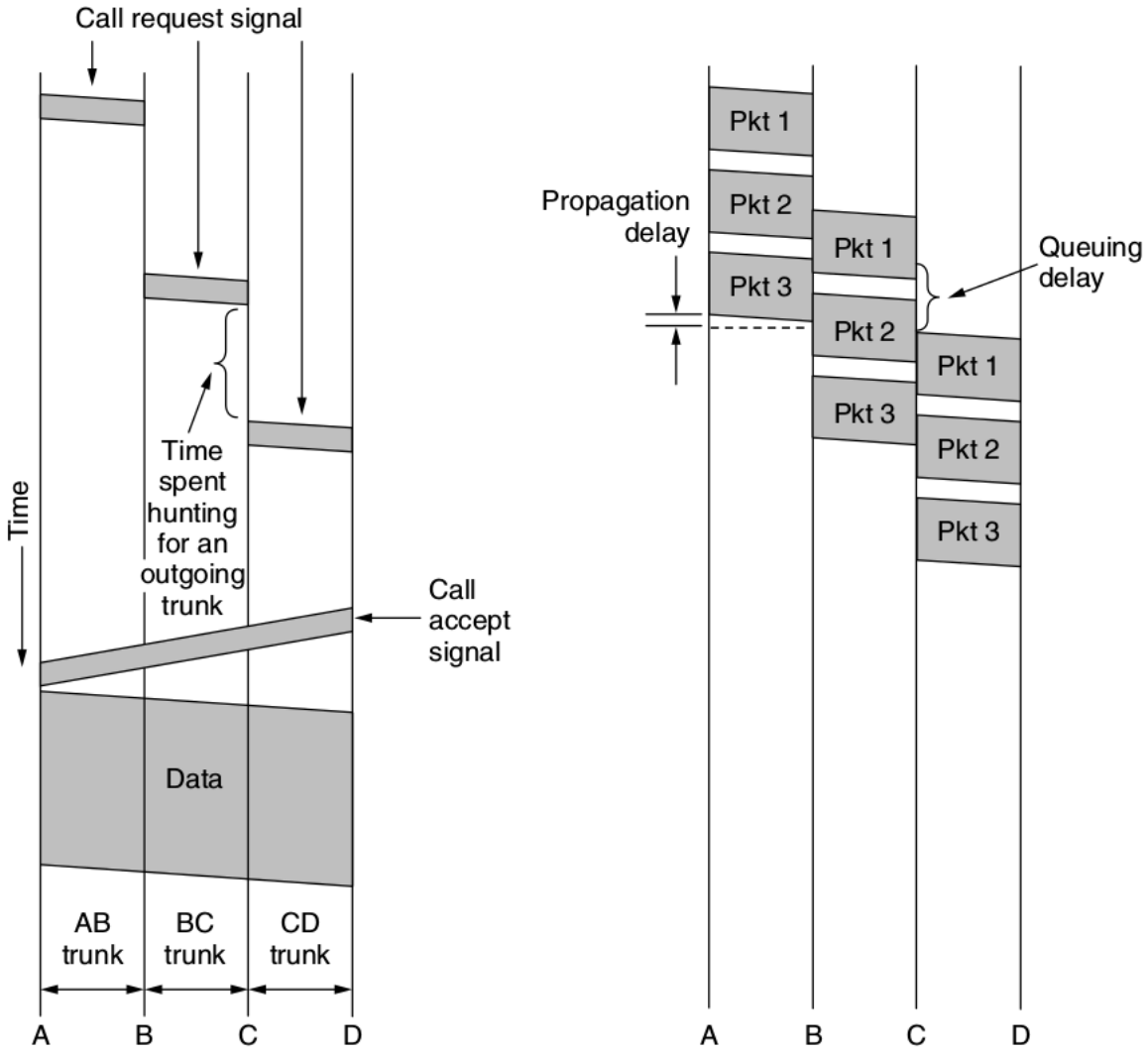
Nachteile

- Aufteilung der Nachricht erfordert Duplizieren der Header-Informationen
- Reihenfolgeprobleme
- Geforderte Dienstgüte schwierig zu garantieren.

Vorteile

- Begrenzung der Paketgröße kann sicherstellen, dass kein einzelnes Paket eine Leitung zu lange belegt.
- Pfade müssen nicht vor der Datenübertragung eingerichtet werden

Leitungs- vs Paketvermittlung



Paket- vs. Leitungsvermittlung

Merkmal	Leitungsvermittlung	Paketvermittlung
Verbindungsaufbau	✓	nicht notwendig
Dedizierter physischer Pfad	✓	✗
Fixe Route der Daten	✓	✗
Reihenfolge der Daten sichergestellt	✓	✗
Fehlertoleranz	✗	✓
Bandbreite-Reservierung	Fix	Dynamisch
Möglicher Datenstau	Bei Verbindungsaufbau	Jederzeit bei der Übertragung eines Pakets
Potenziell verschwendete Bandbreite	✓	✗
„Store and Forward“ Mechanismus	✗	✓

Kapitel 4.4
Wegewahl (Routing)

Internetworking Unit, Routingtabellen, Wegkosten, Quelle-Senken-Baum, Routingverfahren

Begriffe

- Routingalgorithmen beschreiben Wegewahlverfahren
- Verfahrensauswahl und -ausprägung hängt ab
  - von der Routingstrategie (engl. Policy)
  - von den Ziel-/Kostenfunktionen
  - Beispiele:
    - geringe Übertragungskosten
    - geringe Übertragungszeiten
    - gute Leitungsauslastung
    - großer Durchsatz

Begriffe

- Routingverfahren möglichst
  - einfach (algorithmische Komplexität, Netz-Overhead)
  - adaptiv (Last, Topologie)
  - robust (bei Fehlern)
  - fair
- Grundlage ist die Routingtabelle

Probleme

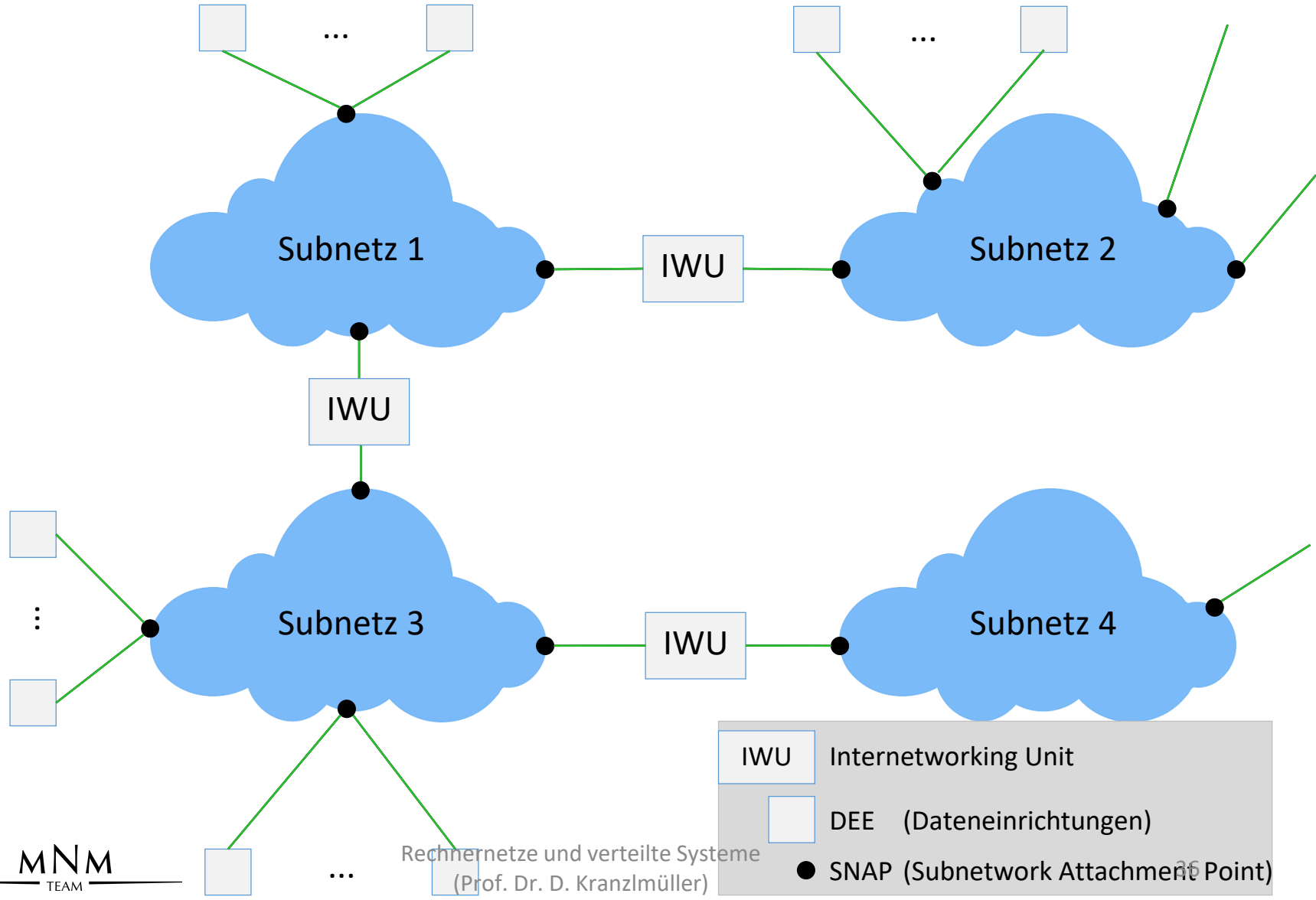
- Zielkonflikte
- Beschreibung der Topologie
  - wie beschrieben (Leitungen, Knoten, Kosten)
  - vollständig / partiell
- Berechnung
  - Wo? (zentral / dezentral)
  - Welche Informationen werden vom Algorithmus benötigt?
  - Wie werden diese Informationen bereitgestellt?
  - Welche Ereignisse stoßen die Berechnung an?
  - Wann wird ein neuer Weg aktiviert?

Verkehrsschild als Basis für Wegewahlentscheidung

Beschriftung ist nur am Standort des Schildes sinnvoll.

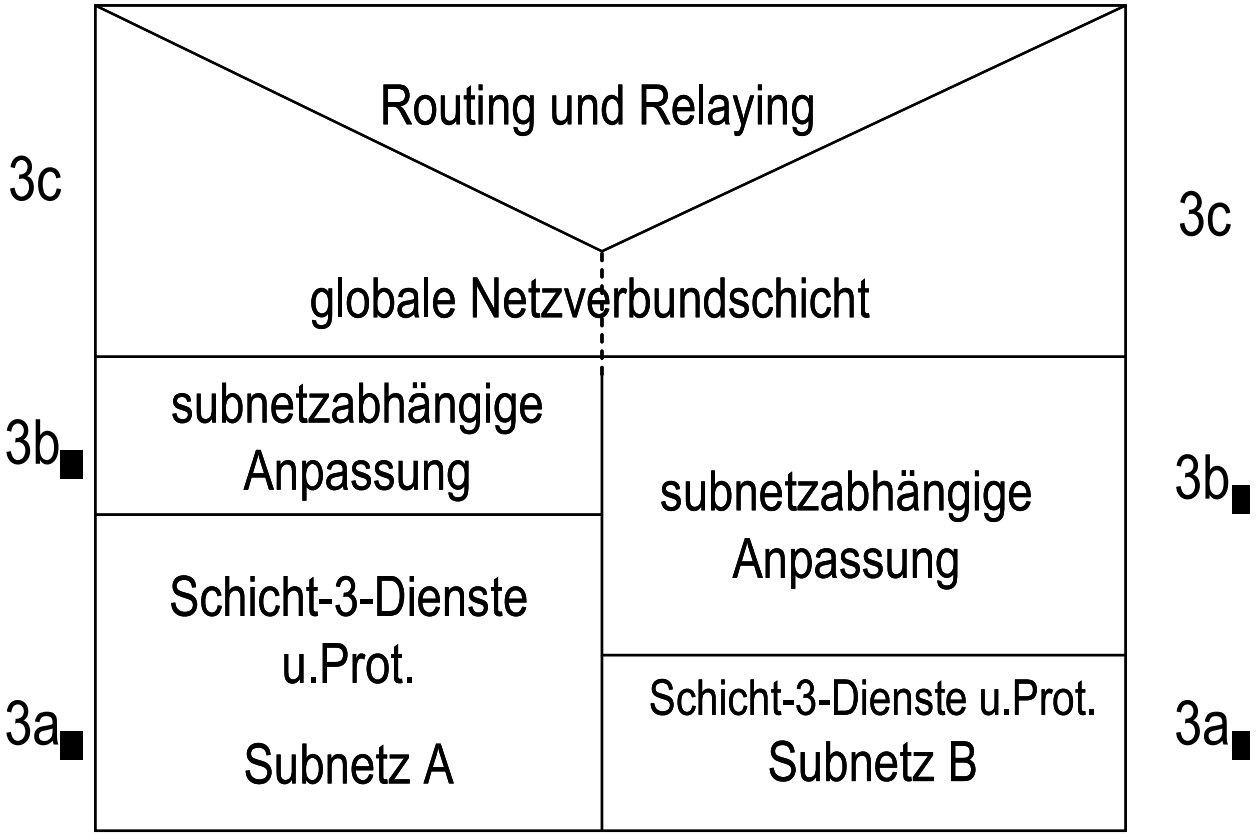
- verschiedene Wege (geradeaus A6, rechts A73)
- indirekt erreichbare Ziele (Regensburg über A3)
- direkt erreichbare Ziele (Flughafen, Messe)

Subnetze (Animation)



Struktur einer IWU (Engl. Inter(net)working Unit)

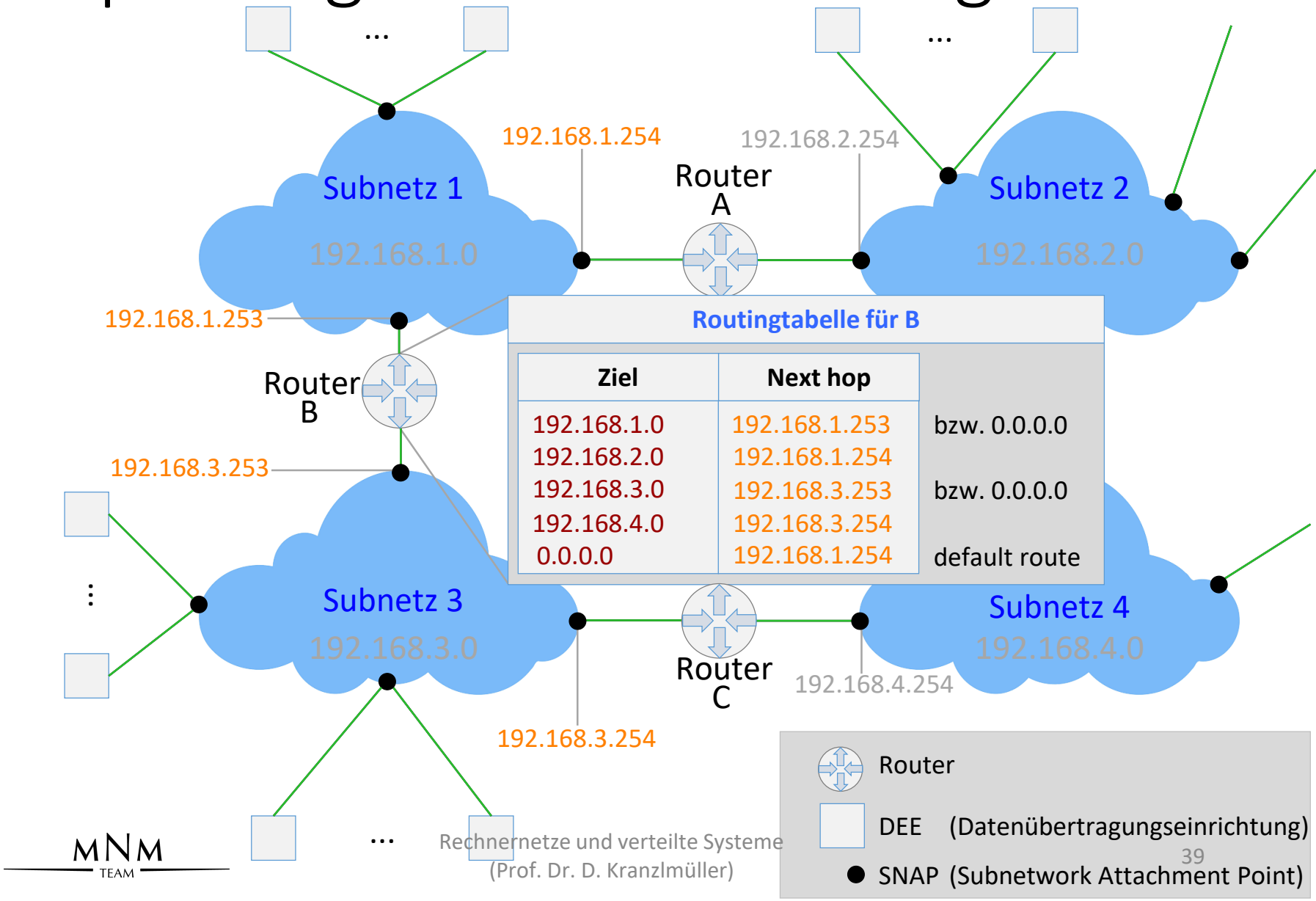
Bsp.: Router/Gateway zwischen Subnetzen A und B



Aufgaben einer IWU (dt. "Kopplungseinheit")

- (Sub-)Netzgrenzen-übergreifende Adressierung/ Adressabbildung
- Anpassung von Nachrichten bzgl. der PDU-Struktur (z.B.: Länge bzw. Größe)
- Abbildung von Diensten (verbindungsorientiert bzw. verbindungslos) und Dienstgüteparametern
- Abbilden von Protokollparametern (z.B. Fenster, Timer)
- Anpassen von Fehlerbehandlungs- und Meldemechanismen
- Globale Wegewahl

Bsp.: Wegewahl mit Routingtabelle



Routingtabellen

- Realisieren Vorgaben zur Nachrichten-Weiterleitung
- Bestimmen die Entscheidungsfindung in IWUs/Routern
- **Default Route**: spezieller Eintrag, wird benutzt, wenn kein anderer Eintrag passt.
- Felder eines Tabelleneintrags (eine Zeile):
  - Ziel (Netz oder Host)
  - Netzmaske (Angabe zur Bestimmung der Netzadresse)
  - Gateway (d.h. Router)
  - Metrik (Kostenwert)
  - Schnittstelle (engl. Interface) (Zielleitung)
- Auf Unix-Derivaten (z.B. Linux): $ route -n

Bsp.: Routingtabelle in einem Host

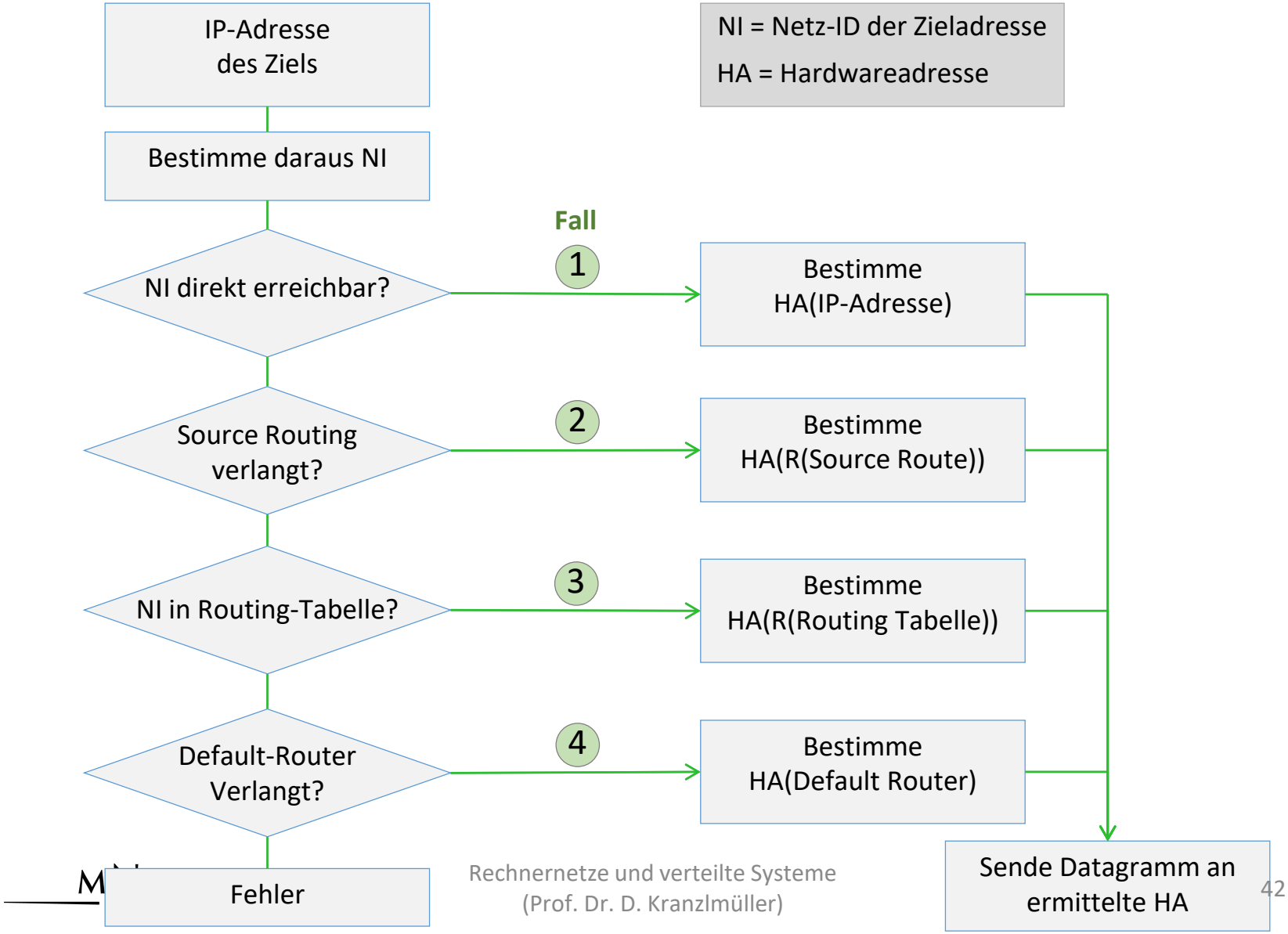
```
danciu@pcheger13:~> /sbin/route -n
Kernel IP routing table
Destination     Gateway         Genmask         Flags Metric Ref    Use Iface
192.168.218.48  0.0.0.0         255.255.255.240 U     0      0        0 eth0
141.84.218.0    0.0.0.0         255.255.255.128 U     0      0        0 eth1
192.168.215.0   192.168.218.126 255.255.255.0   UG    0      0        0 eth1
192.168.217.0   192.168.218.126 255.255.255.0   UG    0      0        0 eth1
192.168.216.0   192.168.218.126 255.255.255.0   UG    0      0        0 eth1
192.168.219.0   192.168.218.126 255.255.255.0   UG    0      0        0 eth1
192.168.235.0   192.168.218.126 255.255.255.0   UG    0      0        0 eth1
192.168.218.0   0.0.0.0         255.255.255.0   U     0      0        0 eth1
192.168.236.0   192.168.218.126 255.255.255.0   UG    0      0        0 eth1
0.0.0.0         141.84.218.126  0.0.0.0         UG    0      0        0 eth1
```

Unterschied Router/Host:

- Router leitet eingehenden Verkehr weiter
- Host leitet lediglich eigenen Verkehr auf die korrekte Schnittstelle

Wie viele Tabelleneinträge braucht ein Host mit einer einzigen Schnittstelle mindestens?

Wegewahlentscheidung im Router



Weiteres

- Wir unterscheiden zwischen:
  - Wegewahl (engl. Route Discovery)
  - Weitergabe (engl. Forwarding)
- Routing ...
  - ist abhängig von Kommunikationsbeziehung
  - meistens Unicast für 1:1 – Beziehung
  - Multicast (Gruppenkommunikation, 1:n, m:n) zunehmend wichtiger (Bsp.: Conferencing, Videoserver, CSCW)
  - Vereinfachungen für spezielle Topologien (Stern, Bus, Ring, Baum, vollständige Vermaschung)
  - Routing für Broadcast muss für NBMA-Netze (engl. Non-Broadcast Multiple Access) emuliert werden

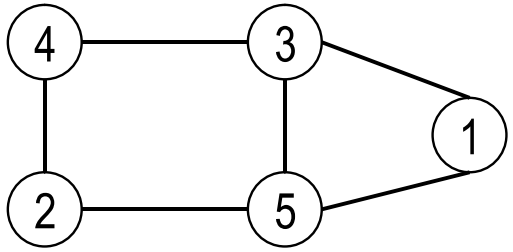
Wegkosten und Optimalitätseigenschaft

- Jedes Wegstück (Kante) in einem Netz (Graph) ist mit gewissen Kosten assoziiert (Kostenfunktion).
- Ein Weg von einer Quelle zu einer Senke ist optimal, gdw. die Summe der Kosten auf diesem Weg minimal ist.
- Optimalitätseigenschaft: Jeder Teilweg auf einem optimalen Weg ist selbst optimal.
- Notation:
  - K = Menge aller Knoten
  - L = Menge aller Kanten,
  - A(j) = Menge der Nachbarn eines Knoten *j*,
  - $W_{qz,f}$: *f*-ter Weg von Quelle *q* zum Ziel *z*,
  - $D_{qz}(W_{qz,f})$: Kosten für f-ten Weg von *q* zum Ziel *z*.

Quelle-Senken-Baum (QSB)

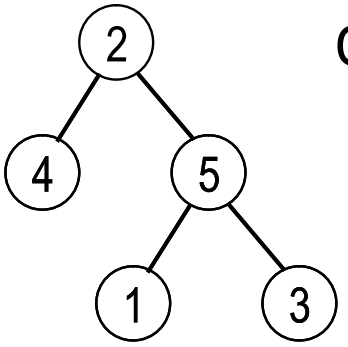
- Ein QSB stellt Wege von einer Quelle (Wurzel des Baums) zu mehreren Senken (Blätter des Baums) dar.
- idR. gibt es zu gegebenen Netzen (mit gegebener Ziel- Quelle und Senken) mehrere mögliche QSB
- Ziel Routingalgorithmen: QSB mit optimalen Wegen finden

Netz

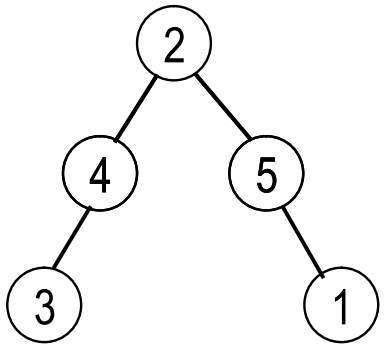


$\forall(j,i):D_{ji}=10$

QSB



oder

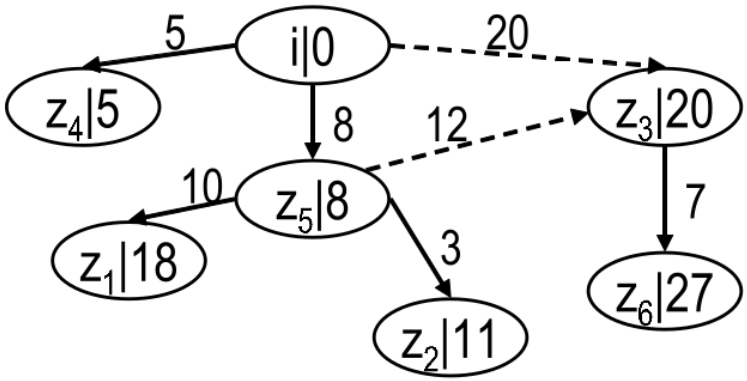


Quelle-Senken-Baum Wegetafel

Wegetafel Knoten i

Ziel	Nachbar	Kosten
z_1	z_5	18
z_2	z_5	11
z_3	z_3, z_5	20
z_4	z_4	5
z_5	z_5	8
z_6	z_3, z_5	27

Quell-Senken-Baum



Shortest Delay First Algorithmus

- Wir bauen den QSB schrittweise von der Wurzel aus auf.
- Alle noch nicht in QSB enthaltenen Wege müssen über Nachbarn von QSB Knoten führen (Kandidatenmenge)
- Aus der Kandidatenmenge H wird der Knoten mit minimalen Kosten ab der Wurzel gewählt.
- Notation
  - A(j), D_{jk}, q wie gehabt
  - $H = \{j \in K, j \in A(k) \mid k \in QSB \wedge j \notin QSB\}$
  - D_{qj}^* : momentane Kosten von q nach j
  - V_j: momentaner Vaterknoten von j zur Erreichung von D_{qj}^*
  - Nächster Knoten aus H liefert Knoten x mit $D_{qx}^* \leq D_{qr}$ für alle $r \in H$

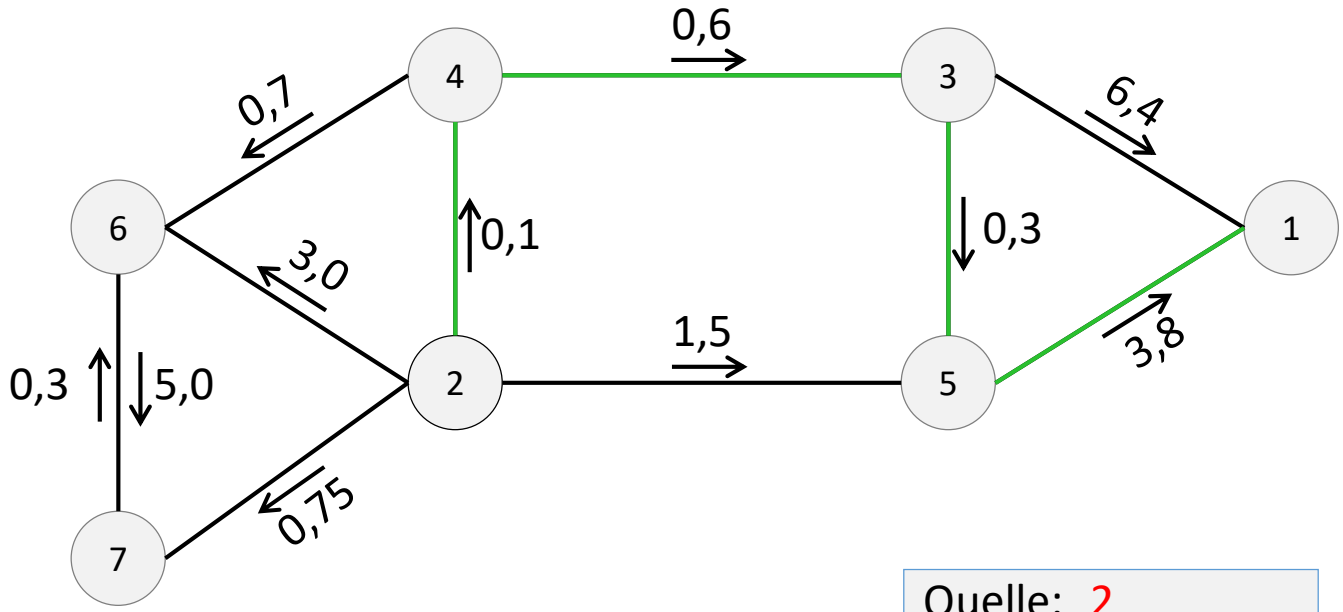
Shortest Delay First Algorithmus Pseudocode

while true **do**
  //Vermessen aller Nachbarn von x
  for all $j \in A(x)$ **do**
    d = $D_{qx}^{opt} + D_{xj}$
    if $j \notin QSB \wedge j \notin H$ **then**
      $H := H \bigcup \{j\}$; $D_{qj}^{*} = d$; $V_j := x$;
    end if
    if $j \in H \wedge d < D_{qj}^{*}$ **then**
      // Es gibt besseren Weg, umhängen
      $D_{qj}^{*} = d$; $V_j := x$;
    else if $H = \{\}$ **then**
      return; *//Ende Algorithmus;*
    end if
  end for
  x := nächster Knoten (H);
  x in QSB aufnehmen als Sohn von V_x und in H löschen;
  $D_{qx}^{opt} := D_{qx}^{*}$;
end while

Wurzel (Quelle)

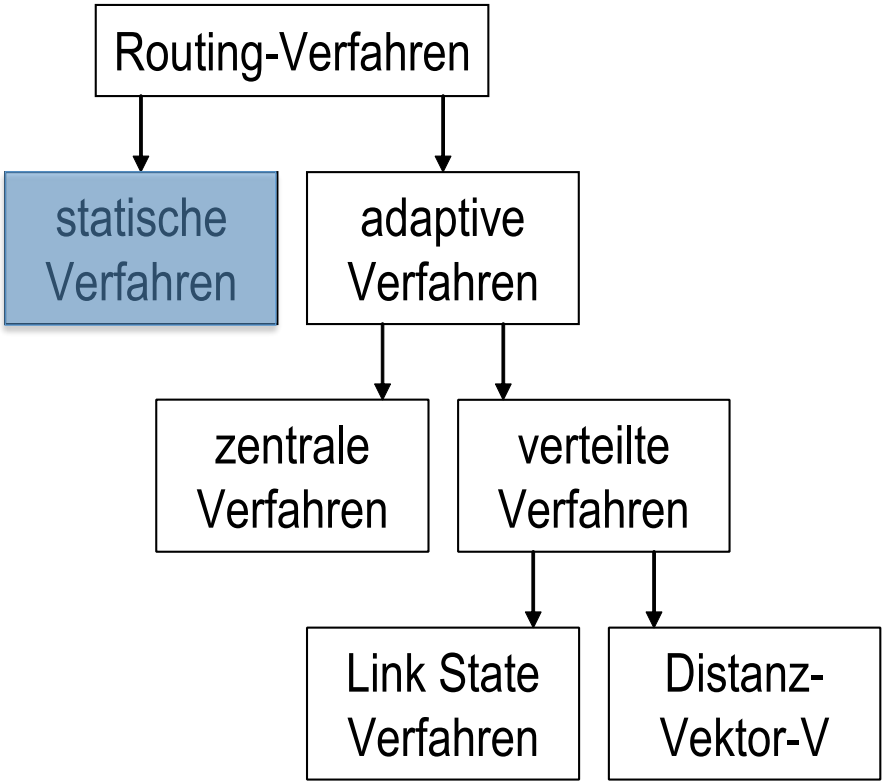
$QSB := q$;
$H := \{\}$;
$D_{qq}^{opt} := 0$
$x := q$
x bezeichnet zuletzt in QSB eingetragenen Knoten

Shortest Delay First Algorithmus Beispiel



Quelle: 2
Aktuell: 4
Senke: 1
Kosten: 0,8 +0,8

Routing Verfahren (Überblick)



Weitere Verfahren:

- isoliertes Verfahren, nicht adaptiv: Flooding
- isoliertes Verfahren, lastabhängig: Hot Potato

Static Routing

Grundidee: Routingtabelle wird einmal auf Basis festgelegter Metriken erstellt (sehr einfach, aber nicht adaptiv).

	Nachbarknoten auf dem Weg von j nach z					
Ziel z	1. Wahl		2. Wahl		3. Wahl	
	Nr.	Gewicht	Nr.	Gewicht	Nr.	Gewicht
	1	0,63	6	0,21	4	0,16

- Berechnen Zufallszahl *x* aus]0,00; 1,00]
  - □ 1. Wahl
  - □ 2. Wahl
  - □ 3. Wahl
- Sonderfälle:
  - nur eine feste Wahl
  - Alternativen nur bei defekter 1. Wahl (engl. Backup Trunk)

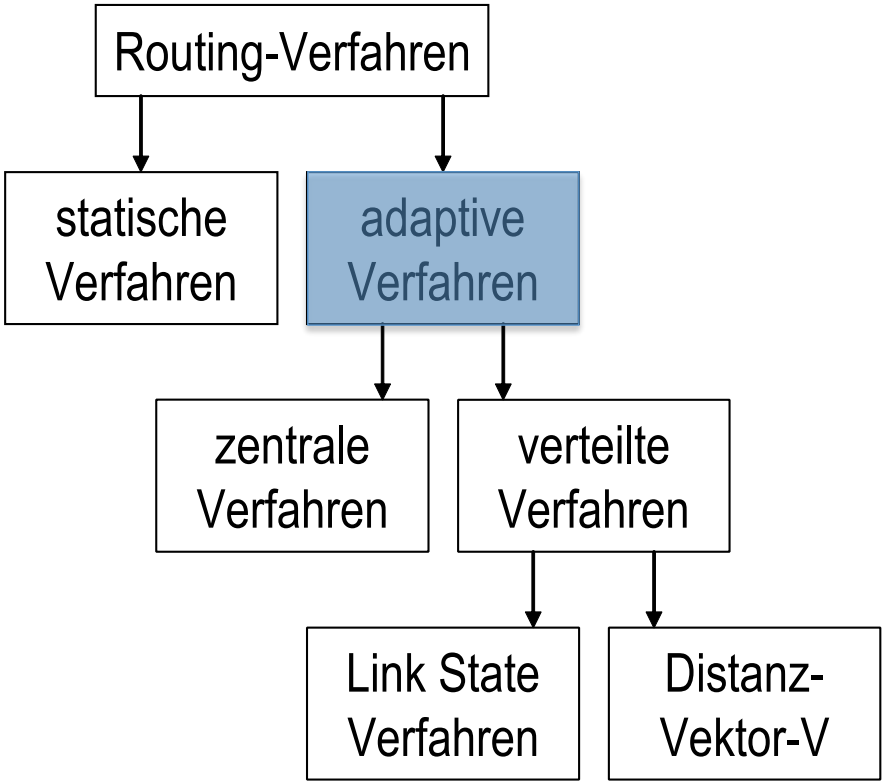
Routing durch Fluten (Engl. Flooding)

- Idee: Jede eintreffende Nachricht wird an alle Nachbarknoten geschickt
- Problem: es entstehen beliebig viele Kopien
- Maßnahmen:
  - nicht zurück an sendenden Knoten
  - Lebensdauerbegrenzung durch Timer oder Hop Count
  - Nachrichtenmerkmal speichern
- Wertung:
  - Verfahren erzeugt Zusatzlast, aber sehr einfach
  - Verfahren extrem robust (z.B. für Militärische Netze)
  - kann für schnelles multiple update von DB benutzt werden
  - enthält immer den optimalen Weg (kann als Metrik gegen andere Verfahren verwendet werden)

Hot Potato

- isoliertes Verfahren
- Ankommende Pakete werden als „hot potato“ (heiße Kartoffel) betrachtet und auf die ausgehende Leitung mit der kürzesten Warteschlange gelegt.
- Eventuell erhebliche Umwege, falls die gewählte Leitung nicht optimal zum Ziel führt
- empfindlich gegen Überlast

Routing Verfahren (Überblick)



Weitere Verfahren:

- isoliertes Verfahren, nicht adaptiv: Flooding
- isoliertes Verfahren, lastabhängig: Hot Potato

Adaptives Verfahren mit Nachbarkenntnis (1/4)

- In allen Knoten wird für jedes Ziel eine Wegetabelle geführt.

z	$a^{opt}(j,z)$	D^{opt}_{jz}

- Jeder Knoten j erhält (periodisch oder ereignisgesteuert) von jedem Nachbarknoten D^{opt}_{kz} für $k \in A(j)$ alle Ziele z, die Kosten des optimalen Wegs $a^{opt}(j,z)$ von j nach z.
- $D_{jz}(W_{kz}) = D_{jk} + D^{opt}_{kz}$ wird für alle $k \in A(j)$ berechnet (D_{jz} lokal, und D^{opt}_{kz} durch Austausch).
- Die Wegewahltabelle wird mit neuem optimalem Nachbarn aktualisiert.

Adaptives Verfahren mit Nachbarkenntnis (2/4)

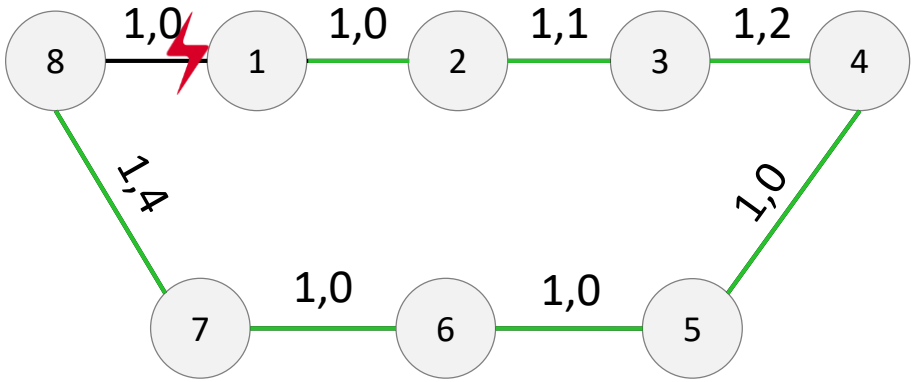
- Beispiel: Distanz-Vektor-Alg. (Bellman-Ford-Alg)
- wird im Internet vom Routing Information Protokoll (RIP) benutzt.
- Merkmale:
  - jeder Knoten hat Routingtabelle (Adresse, Distanz),
  - Update geschieht periodisch
  - jede Kante ist mit Gewicht belegt
  - Distanz ist Summe der Gewichte auf dem Weg zum Ziel
- Initialisierung mit Eintrag
  - dessen Ziel dem lokalen Knoten entspricht,
  - dessen Next-Hop nicht angegeben ist und
  - dessen Distanz auf 0 gesetzt ist

Distanz-Vektor-Algorithmus

```
while true do
    warte auf nächste Routing Nachricht vom Nachbarn N;
    for all Eintrag in Nachricht do
        Z sei Ziel im Eintrag; D sei Distanz zu N;
        Berechne: C := D+ Gewicht der Kante des Eintreffens von N;
        //Prüfe und aktualisiere die lokale Routingtabelle:
        if es gibt keine Route zu Z then
            ergänze Eintrag mit Ziel Z, NextHop N, Distanz C;
        else if es gibt eine Route mit NextHop N then
            ersetze Distanz in existierender Route durch C;
        else if es gibt eine Route mit D > C then
            ändere NextHop auf N und Distanz auf C;
        end if
    end for
end while
```

Adaptives Verfahren mit Nachbarkenntnis (3/4)

- Verfahren sehr einfach, geringer Rechenaufwand
- Topologie muss nur partiell bekannt sein
- Zusätzliche Netzbelastung
- Wegewahlinformation breitet sich nur langsam aus ➔ kann zu Inkonsistenz führen



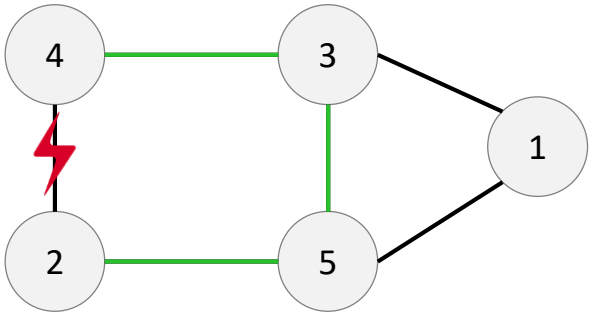
Aufgabe 1:

Berechne die ersten Schritte zu den Zeitpunkten $t_1,\ldots,t_8$, wenn zwischen t_1 und t_2 das Wegstück $W_{1,8}$ ausfällt.

Adaptives Verfahren mit Nachbarkenntnis (4/4)

Aufgabe 2:

Betrachte die Wegetabellen in Knoten 3 unter der Annahme, dass das Wegstück $L_{2,4}$ während des Intervalls $[t_x,t_{x+1}[$ ausfällt und während des Intervalls $[t_y,t_{y+1}[$ wieder repariert wird.



Gemeinsam adaptive Wegberechnung (1/2)

Ziel

- möglichst alle Knoten sollen aktuelle Wegewahlinformationen haben.
- langsame Ausbreitung von Anpassungsereignissen soll vermieden werden.

Verfahren (auf Basis des SDF-Algorithmus)

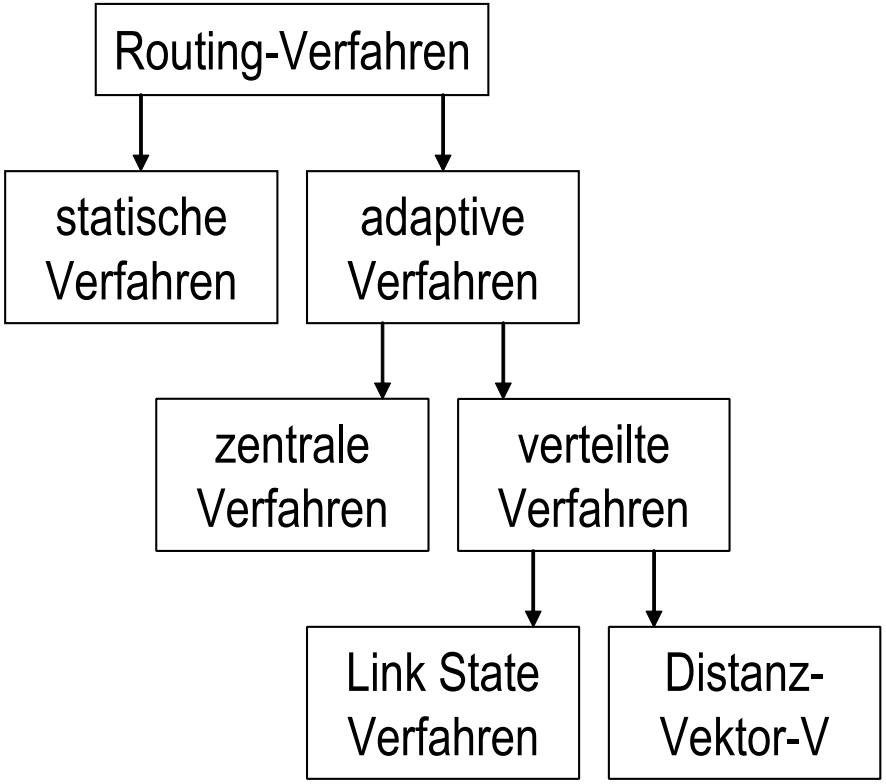
- jeder Knoten kennt Topologie und Bewertung
- jeder Knoten erhält von jedem anderen Knoten für alle dessen lokalen Verzögerungsvektor .
- Verteilung der Vektoren z.B. durch Fluten

Gemeinsam adaptive Wegberechnung (2/2)

Beispiel: Link-State-Routing, SPF-Routing

- wird im Internet im Routingverfahren OSPF verwendet
- jeder Router versucht seine Nachbarn kennenzulernen
- jeder Router bildet ein *Link State Packet* (LSP) mit Namen der Nachbarn und Gewichten der zugehörigen *Links*
- Die LSP werden an alle Rechner geschickt, jeder Rechner speichert die zuletzt erhaltenen LSP aller anderen Router. Jeder Router kennt die vollständige Netztopologie.
- QSB wird nach SDF-Verfahren (Dijkstra) berechnet.

Routing Verfahren (Wiederholung)

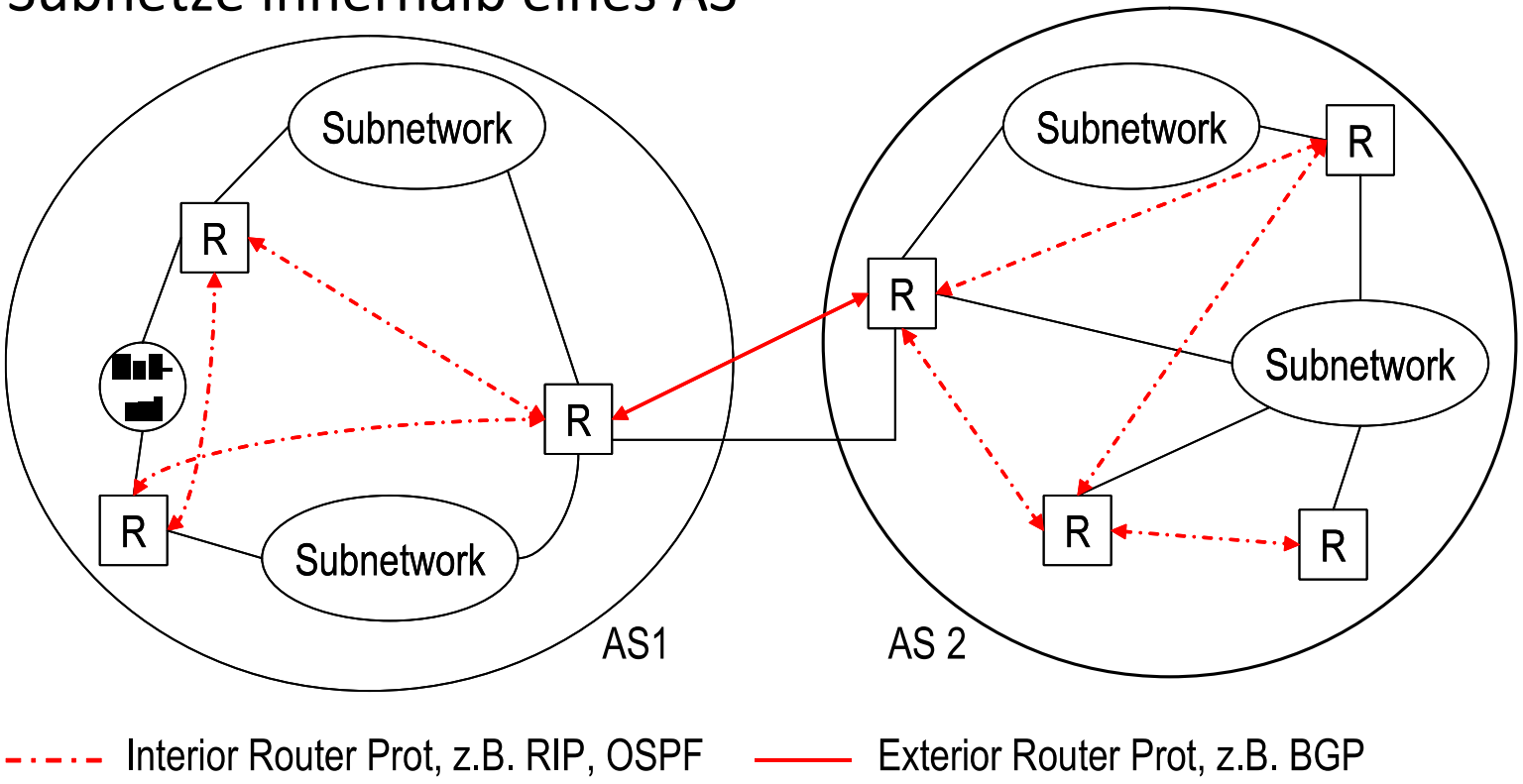


Routing im Internet: Autonome Systeme (1/2)

- Ein AS ist eine Menge von Routern unter der Kontrolle einer administrativen Domäne (Routing Domäne).
- AS sind genehmigungspflichtig (durch die IANA/ICANN) und haben eine eindeutige numerische Kennzeichnung (Bsp.: das LRZ/MWN ist das AS 12816)
- Im Innern von AS wird ein IGP (engl. Interior Gateway Protocol) verwendet.
- Zwischen AS wird ein EGP (engl. Exterior Gateway Protocol) verwendet.
- Abkommen zwischen AS (ggf. mit Zahlungen verbunden)
  - Peering: Routen in/aus Netz eines fremden AS
  - Transit: Routen durch/über das Netz eines fremden AS
  - Technische Realisierung durch Network Exchange

Routing im Internet: Autonome Systeme (2/2)

- Autonome Systeme (AS): administrativ selbstständige Netze
- Subnetze innerhalb eines AS



Gateway Protokolle (Überblick)

IGP (engl. Interior Gateway Protocol) (intra-AS)

- RIP (engl. Routing Information Protocol): Distanz-Vektor-Algorithmus
- OSPF (engl. Open Shortest Path First): Link-State-Routing

EGP (engl. Exterior Gateway Protocol) (inter-AS)

- BGP (engl. Border Gateway Protocol): ähnlich dem Distanz-Vektor-Verfahren

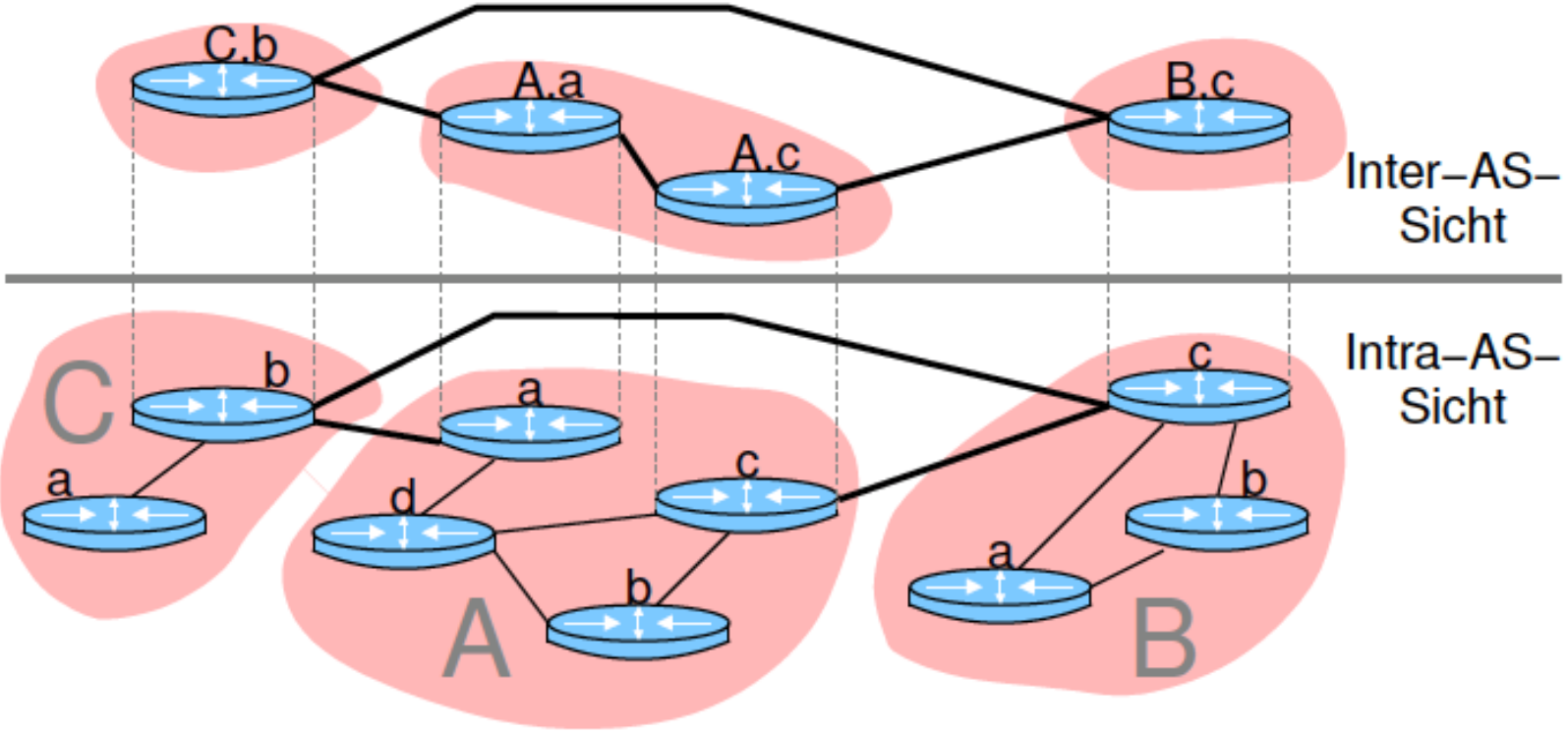
Border Gateway Protocol

EGP-Kriterien für Wegewahl: finanziell, organisatorisch, eher nicht-technisch

BGP: Arbeitsweise ähnlich Distanz-Vektor-Verfahren

- Angabe des Pfades als Sequenz der zu durchlaufenden AS (anstatt Distanzangabe)
- keine Metrik: Pfade realisieren policy-basiertes Routing indem bestimmte Pfade ausgeschlossen oder uninteressant gemacht werden können.
- Aufgaben: Neighbour acquisition, neighbour reachability, network reachability
- Session-/TCP-basiert (Port 179): Es besteht eine ständige Verbindung zwischen „BGP-speakers“.
- In den RFCs 1772-1774 spezifiziert (CIDR-Unterstützung im RFC 4271)

Routing im Internet: Intra-AS-Routing



Fragen zu Wegewahl/Routing

- Warum war es sinnvoll, im Internet RIP durch OSPF abzulösen?
- Was ist der Kernalgorithmus bei Link State Routing?
- In welchen Fällen ist Flooding ein sinnvolles Verfahren?
- Nennen Sie Kostenfunktionen, die dem Optimalitätskriterium genügen?
- Was versteht man unter autonomen Systemen?
- Neuere BGP-Versionen unterstützen CIDR. Was bedeutet diese Aussage?

Inhalt von Kapitel 4

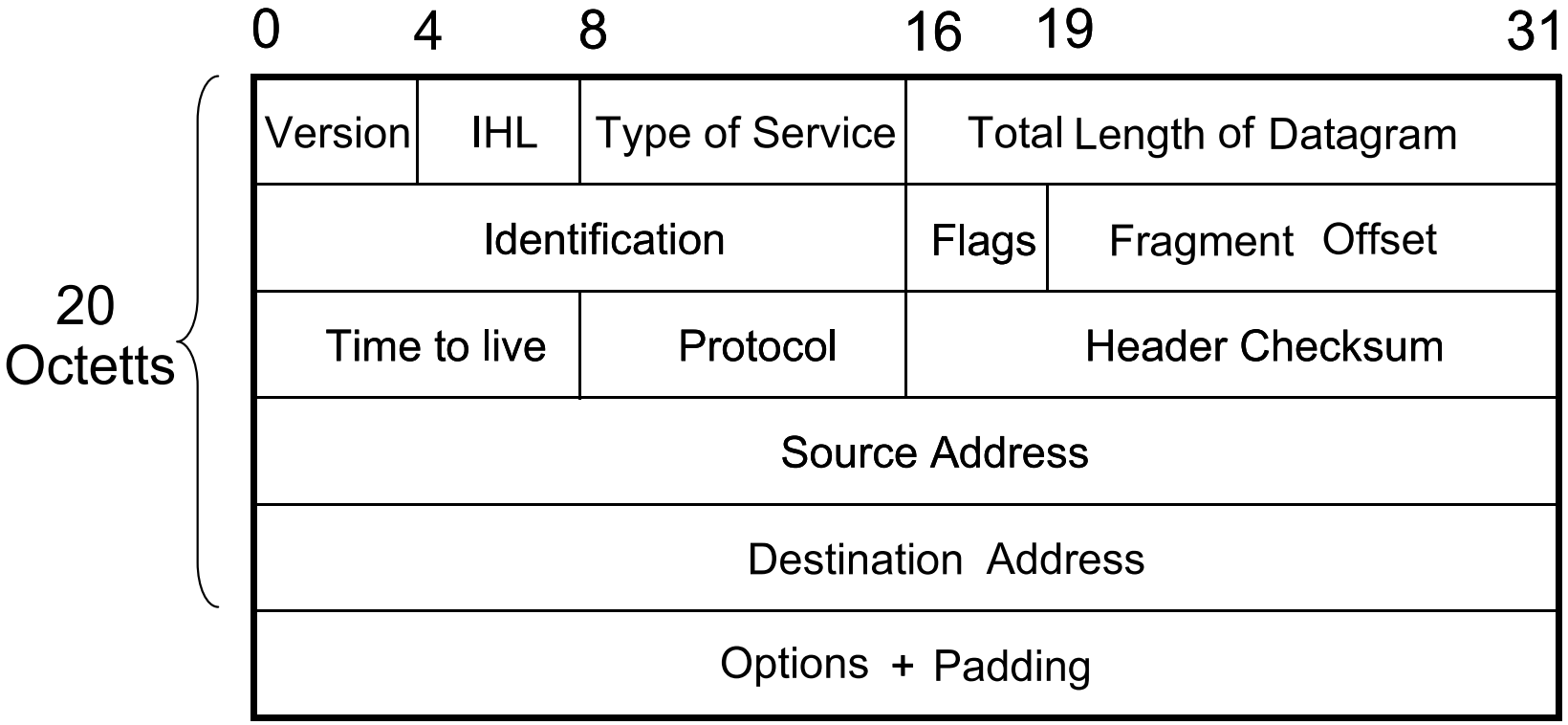
1. Ziele und Rahmenbedingungen
2. Wegewahl auf anderen Schichten
3. Vermittlung (Pfadschaltung)
4. Wegewahl (Routing)
5. Internet Protokoll (v4)
6. Internet Protokoll (v6)

Kapitel 4.5
Internet Protokoll (v4)

Überblick IP

- Verbindungsloser Dienst mit zwei Dienstprimitiven (SEND, DELIVER)
- Netzglobale Adressierung mit Hilfe von IP-Adressen
  → Kapitel 2: Namen und Adressen
- Protocol: ID des IP-Nutzers (SAP-Adresse)
  Bsp.: 1 ICMP, 6 TCP, 9 UDP, 46 RSVP
- Unterstützt Fragmentierung und Reassembly
  (zum Umgang mit MTUs)
- Time-to-Live (TTL), gemessen in „Hops“ (Wegstücken)
- IP-Datagramm bestehend aus Header- und Datenteil

Der IP-Header (PCI) als Grafik



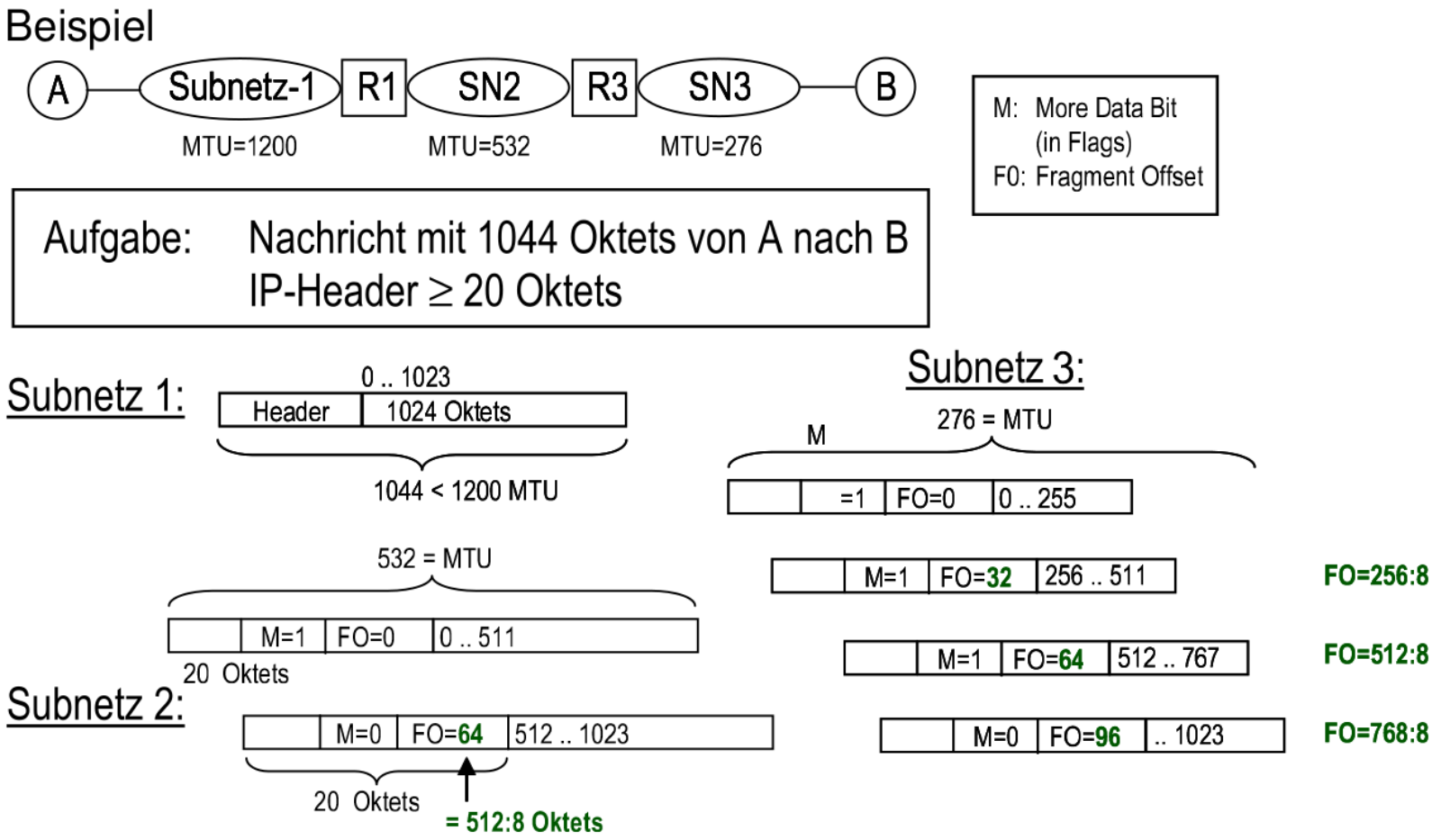
Der IP-Header (PCI) (1/2)

- **Version** (4 Bit): Bei IPv4 immer 4.
- **IHL (Internet Header Length)** (4 Bit): Header Länge in 32-Bit-Wörtern (mindestens 5, wenn keine Optionen verwendet werden).
- **Type of Service** (8 Bit): Dienstklasse/Priorität und ECN (engl. Explicit Congestion Notification).
- **Datagramm Länge** (16 Bit): Gesamtlänge des Datagramms in Bytes (inkl. Header): max. 64 Kbyte (65545 Oktetten)
- **Identifikation** (16 Bit): Dient der Defragmentierung: alle Fragmente eines zusammengehörenden Pakets enthalten die selbe Identifikation.
- **Flags** (je 1 Bit)**:**
  - 1 Unbenutztes Bit („evil Bit“)
  - DF: *Don‘t Fragment* (Dient der MTU Ermittlung)
  - MF: *More Fragments* (Es fehlen noch weitere Paketfragmente)

Der IP-Header (PCI) (2/2)

- **Fragment Offset** (13 Bit): Die Stelle, die ein Fragment im originalen Paket inne hält, gemessen in 8-Byte Blöcken.
- **Time to Live** (8 Bit): Anzahl hops, die das Datagramm machen darf bevor es verworfen wird.
- **Protocol** (8 Bit): ID des IP-Nutzers (für welchen SAP ist das Paket bestimmt?)
- **Prüfsumme** (16 Bit): Einerkomplement der Summe der 16-Bit-Halbwörter des Headers. (Muss nach jeder Teilstrecke neuberechnet werden, da sich die TTL ändert.)
- **Quell-/Zieladresse** (je 32 Bit): IP-Adressen der Netzschnittstellen von Quelle und Ziel.
- **Optionen** (n * 32 Bit): Wählbare Eigenschaften, wie z.B. ein Timestamp.

IP-Fragmentierung



IP und unterstützende Protokolle

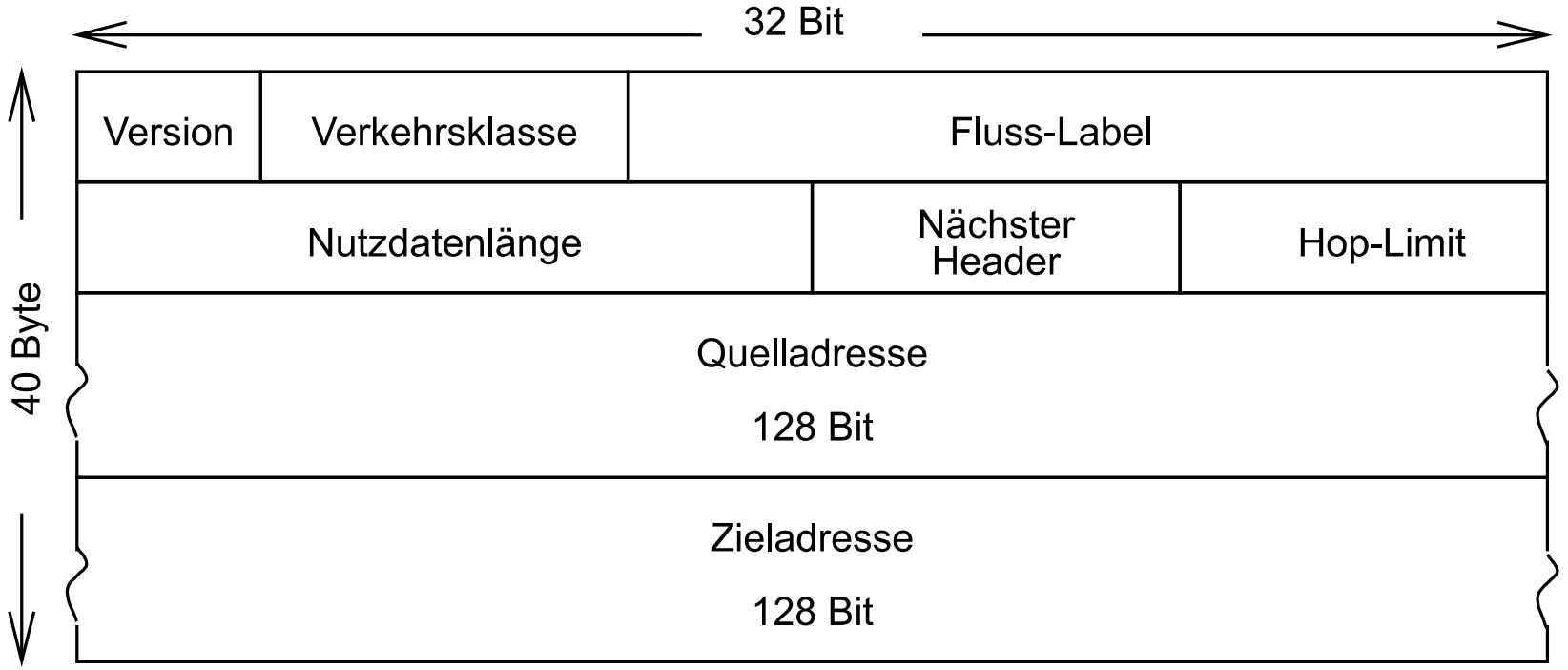
- ICMP (engl. Internet Control Message Protocol),
  - RFC 792, 1122
  - Benachrichtigungsprotokoll, dient dem Austausch von Informationen und Fehlermeldungen über IP.
  - Bsp.: Destination Unreachable, Time Exceeded, Redirect, usw.
  - Basis für traceroute
- DHCP (engl. Dynamic Host Configuration Protocol)
  - Dynamische temporäre Zuweisung von IP-Adressen
- DNS (engl. Domain Name System)
  - Abbildung von Namen auf IP-Adressen
  - Interaktion zwischen Resolver (Client) und Server (Directory)
- ARP (engl. Address Resolution Protocol), RFC 826
  - Abbildung von IP-Adressen auf MAC-Adressen (Schicht 2)
- IPsec (engl. Internet Protocol Security):
  - Authentifizierung und Verschlüsselung von IP-Paketen

Kapitel 4.6
Internet Protokoll (v6)

Überblick

- Motivation
  - Adressknappheit bei IPv4 bei wachsendem Bedarf an Adressen.
  - 1. Februar 2011: zentraler Vorrat an IPv4-Adressen erschöpft.
- Neue und ausgebaute Funktionalität (im Vgl. zu IPv4)
  - erweiterter Adressraum (128 statt 32 bit)
  - dynamische Adresszuweisung (Mobilitätsunterstützung)
  - Erweiterungen für QoS auch bzgl. Streams wie Sprache, Video teilweise verbindungsorientierte Protokollaspekte
  - Ressource Allocation (Reservierungsprotokolle)
  - eingebaute Sicherheitsmechanismen (IPsec)
  - Router Anweisungen (hop-by-hop-options)
- Status
  - IPv6 wächst; In gängigen Betriebssystemen implementiert.
  - China, Japan sind Vorreiter (unter anderem, weil sie von der IPv4 Adressknappheit besonders betroffen sind).

Der IPv6-Header (PCI) als Grafik

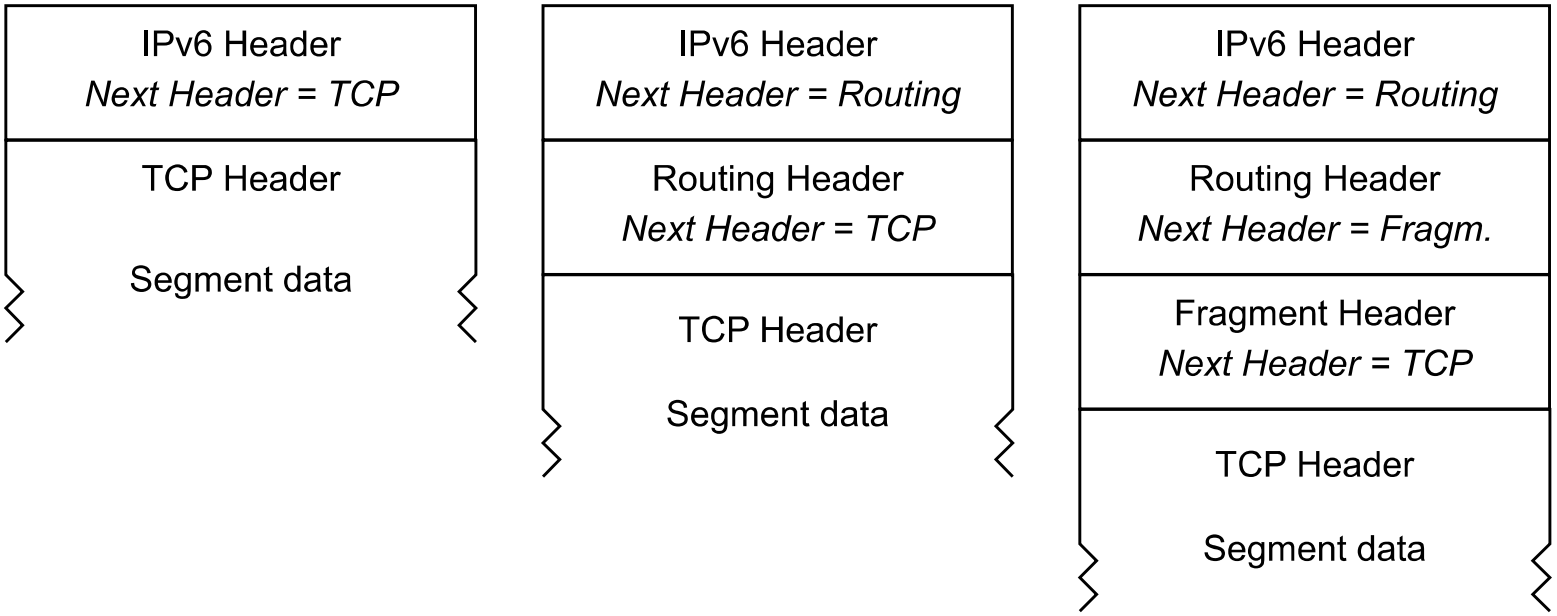


- Der IPv6-Header hat eine feste Größe von 40 Byte.
- 32 dieser Bytes gehen allein an die Quell- und Zieladresse.
- Wenn das nächste Header Feld entsprechend gesetzt ist, können Erweiterungs-Header angehängt werden.

Der IPv6-Header (PCI)

- **Version** (4 Bit): Bei IPv6 immer 6.
- **Verkehrsklasse** (8 Bit): Entspricht dem „Type of Service“ Feld des IPv4 Headers.
- **Fluss Label** (20 Bit): Paket kann als teil eines zusammengehörigen Paketstroms behandelt werden Verbindungsorientierter Protokollaspekt.
- **Nutzdatenlänge** (16 Bit): Anzahl der Nutzdatenbytes im Paket.
- **Nächster Header** (8 Bit): Typ des auf den Header folgenden Erweiterungs-Headers. Wenn keine weiteren Header folgen, entspricht dieses Feld dem Protokollfeld des IPv4 Headers.
- **Hop-Limit** (8 Bit): Entspricht dem Time to Live Feld des IPv4 Headers (der Name wurde geändert um die tatsächliche Nutzung besser widerzuspiegeln).
- **Quell-/Zieladresse** (je 128 Bit): IPv6-Adressen der Netzschnittstellen von Quelle und Ziel.

Erweiterungs-Header

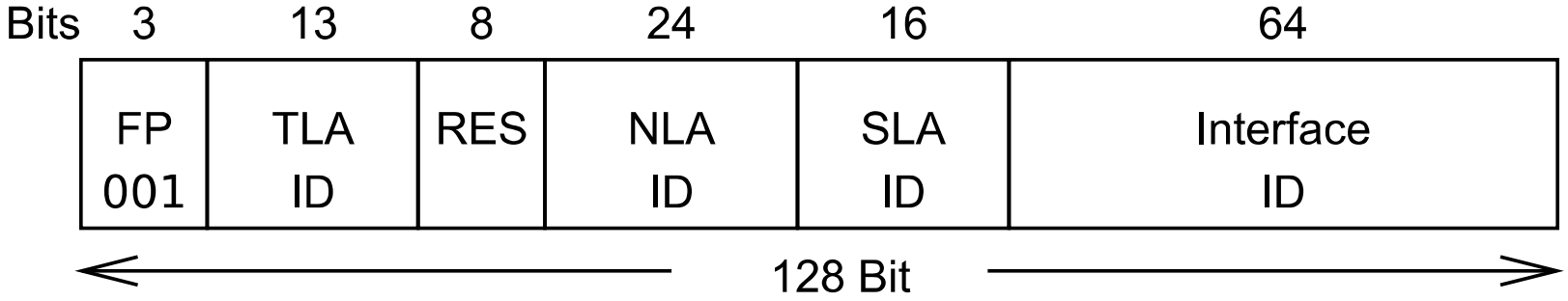


- Original 6 Header-Typen als Teil der IPv6 Spezifikation (RFC 2460)
  Definiert: Hop-by-Hop Options, Routing, Fragment, Destination Options, Authentication, Encapsulating Security Payload
- Weitere Header-Typen (in der Zukunft) möglich.
- Anzahl (i.d.R. eins) und Reihenfolge der Header können zur effizienten Verarbeitung vorgegeben werden.

IPv6-Adressen

- 128-Bit Unicast-, Multicast- und Anycast-Adressen
  - Unicast: Übertragung von einer Quelle zu einem Ziel
  - Multicast: für Mehrpunktverbindungen
  - Anycast: Adressierung eines beliebigen Hosts einer Gruppe (zur Lastverteilung und Ausfallsicherheit)
- Notation
  - Acht, durch Doppelpunkte getrennte, 16 Bit Hexadezimalzahlen. Bsp.: 1080:0:0:0:8:800:200C:417A
  - Einmal pro Adresse dürfen mehrere Nullwertige 16-Bit-Gruppen durch „::“ zusammengefasst werden: 1080::8:800:200C:417A
  - Präfixschreibweise wie CIDR: 12AB::CD30:0:0:0:0/60
- Sonderadressen
  - Nicht angegeben (Abwesenheit einer Adresse): 0:0:0:0:0:0:0:0
  - Loopback Adresse: 0:0:0:0:0:0:0:1 oder ::1

Globale Unicast-Adressen

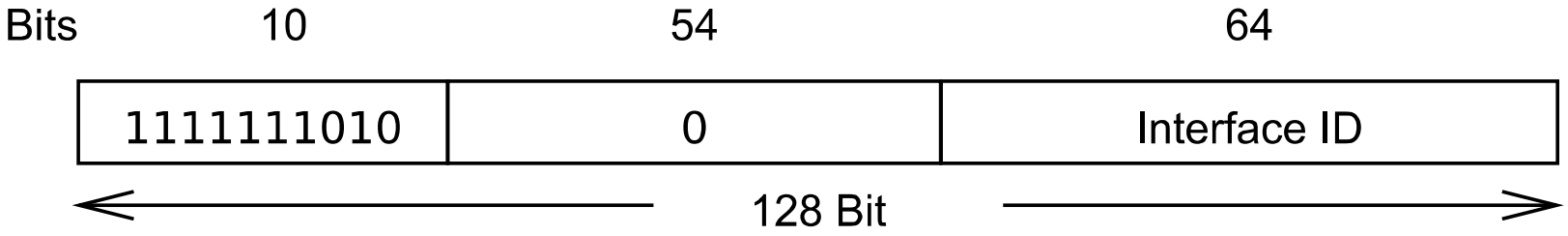


- Adress-Felder:
  - **Format Prefix (FP), Top-Level Aggregation ID (TLA ID), RES, Next-Level Aggregation ID (NLA ID):** Ergeben zusammen den globalen Routingprefix (eine Art strukturierte Netz ID).
  - **Site-Level ID (SLA ID):** Entspricht einer Subnetz ID.
  - **Interface ID**: Adressiert die Netzschnittstelle auf dem Host
- Dreistufige Adresshierarchie

Lokale Unicast-Adressen

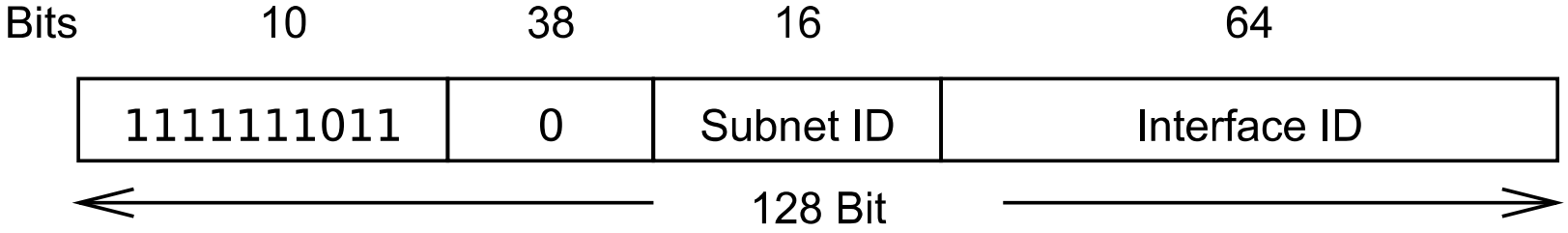
Link-local

- Bezug nur auf einen bestimmten Link
- Einsatz z.B. für Autokonfiguration



Site-local

- Geltungsbereich umfasst einen Standort/eine Site
- ähnlich privater Adressblöcke bei IPv4
- seit 2004 abgeschafft („deprecated“), RFC 3879

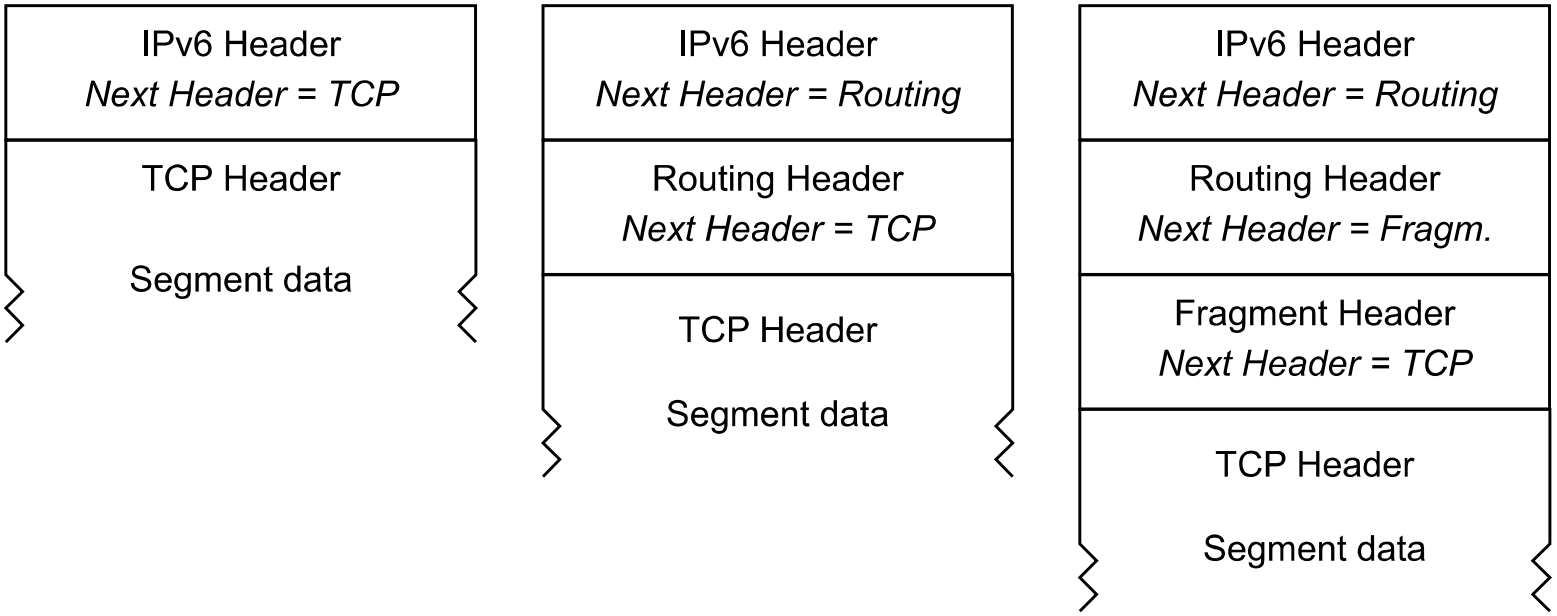


IPv6-Adressen auf Basis von Schicht-2-Adressen

- Idee: Nutzung gerätegebundener Adressen bei der Bildung von IPv6-Adressen
- Anwendung nur auf Interface Identifier (64 Bits)
- Extended Unique Identifier (EUI-64) (Definition der IEEE)
- z.B. anwendbar mit Ethernet, Token Ring, FDDI . . .
- Beispiel: IEEE 802, 48 Bit auf 64 Bits Abbilden

  00-23 Kennung des Herstellers mit Flag-Bits
  6 universal/local (1 = globaler Geltungsbereich)
  7 individual/group
  24-39 1111111 111111110
  40-63 spezifisch für Geräteinstanz

Erweiterungs Header



- Original 6 Header-Typen als Teil der IPv6 Spezifikation (RFC 2460)
  Definiert: Hop-by-Hop Options, Routing, Fragment, Destination Options, Authentication, Encapsulating Security Payload
- Weitere Header-Typen (in der Zukunft) möglich.
- Anzahl (i.d.R. eins) und Reihenfolge der Header können zur effizienten Verarbeitung vorgegeben werden.

Extension Header - Beispiele

- Hop-by-Hop Options Header
- Routing Header
- Fragment Header

Der Hop-by-Hop Options Header

- Zweck: Angabe von Optionen, die jeder Knoten auf dem Weg berücksichtigen soll. (z.B. Padding)
- Felder:
  - **Next Header** (8 Bit): Angabe des folgenden Headers (wie bekannt)
  - **HEL (Header Extension Length)** (8 Bit): Länge des Erweiterungs-headers in Bytes.
  - **Optionen:** als (Typ, Länge, Wert) Tupel Codiert:
    - **Typ** (8 Bit): Art der Option
    - **Länge** (8 Bit): Länge des Wertes der Option in Byte
    - **Wert** (bis zu 255 Byte): beliebige Informationen
- Bisherige Optionen:
  - Datagramme mit einer Länge größer als 64 KB.

Der Routing Header

- Zweck: Angabe von Zwischenstationen die das Paket auf seinem Weg durch das Netz der Reihe nach durchlaufen muss.
- Felder
  - **Next Header, Header Extension Length**: wie bekannt
  - **Routing Type**: identifiziert Routing-Header Variante (default: 0)
  - **Segments Left**: Anzahl verbliebener Zwischenstationen, wenn 0 erreicht ist, wird der Routing Header ignoriert.
  - **Typspezifische Daten:** abhängig von Routing Type.

Fragmentierung

- Anders als bei IPv4, wird die Fragmentierung von IPv6 Paketen ausschließlich in den Endsystemen und nicht im Transitsystem durchgeführt.
- Nicht-Fragmentierbarer Teil: IPv6-Header + alle Header mit Ende-zu-Ende Relevanz
- Aufbau von Fragment-Paketen:
  - Nicht-Fragmentierbarer Teil
  - Fragment Header (nächste Folie)
  - Inhalt des ersten Fragments

Der Fragment Header

- Zweck: Senden eines IP-Pakets, das größer ist als die MTU (Maximum Transmissible Unit)
- Felder:
  - **Next Header, Header Extension Length**: wie bekannt
  - **Fragment Offset:** Offset der Daten, die diesem Header folgen
  - **MF Flag (More Fragments)**: wie bei IPv4 (0 bedeutet letztes Fragment, 1 bedeutet weitere Fragmente folgen.)
  - **Identification:** eindeutige Kennung des IP-Pakets, das fragmentiert wird.

IPv4 zu IPv6 Migration

- Betriebliche Aufgabenstellung
  - Möglichst kosteneffiziente Ersetzung der IPv4-Technologie.
  - Koexistenz von IPv4 und IPv6 während der Übergangsphase.
- Prinzipielle Ansätze
  - Dual-Stack: Knoten besitzen sowohl IPv4- als auch IPv6-Implementierung
  - Tunneling (IPv6-Verkehr über IPv4-Netz und umgekehrt)
    - Kapselung von Paketen im tunnelnden Protokoll
    - Managment des Tunnels: Erstellung, Umgang mit Fragmenten
- Übersetzung
  - Protokollübersetzung in Übergangsknoten/Router
  - in Programmbibliotheken (APIs zur Netzprogrammierung)